

FINAL REPORT

HIGH SPEED, HIGH ACCURACY STAGE FOR ADVANCED LITHOGRAPHY

December 2001

SPONSORED BY

DEFENSE ADVANCED RESEARCH PROJECTS AGENCY

Issued by
U.S. Army Aviation Missile Command

Under
Contract No. DAAH01-99-C-R080

OPTRA, Inc
461 Boston Street
Topsfield, MA 01983
(978) 887-6600

Principal Investigator: Michael Hercher

Program Manager: David Mansur

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either express or implied, of the Defense Advanced Research Projects Agency or the U.S. Government.

Approved for public release; distribution unlimited

20020107 213

OPTRA

December 28, 2001

Commander
U.S. Army Aviation and Missile Command
ATTN: AMSAM-RD-WS-DP-SB/Mr. Roach
Bldg 7804, Room 221
Redstone Arsenal, AL 35898-5248

RE: Contract No. DAAH01-99-C-R080
SUBJ: Final Technical Report

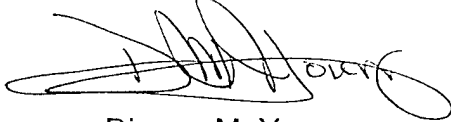
Dear Mr. Roach:

As required by the above referenced contract, I am forwarding two (2) copies of the above referenced report. Also enclosed are two (2) copies of the required DD250. Please sign and return one copy to me for our records.

Please feel free to contact me should you have any questions of an administrative nature, or Mr. Michael Hercher, Principal Investigator, for questions of a technical nature.

Sincerely,

OPTRA, Inc.



Dianne M. Young
Contracts Administrator

Enclosure

Cc: AMSAM-RD-OB-R (1 copy)
AMSAM-RD-WS (1 copy)
Dr. D. Patterson (1 copy)
ASBD/SBIR (1 copy)
ASBD/DARPA Library (1 copy)
DTIC (2 cps)
DCMC/GFOG (1 copy DD250)

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project, Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE 12/28/01	3. REPORT TYPE AND DATES COVERED Final 2/4/99 - 12/31/01
4. TITLE AND SUBTITLE A High Speed, High Accuract Stage for Advanced Lithography		5. FUNDING NUMBERS DAAH01-99-C-R080
6. AUTHOR(S) Michael Hercher and David Mansur		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(S) OPTRA, Inc 461 Boston Street Topsfield, MA 01983		8. PERFORMING ORGANIZATION REPORT UNMBER 298
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESSES(S) U.S. Army Aviation and Missile Command AMSAM-RD-WS-DP-SB Bldg 7804, Room 221 Redstone Arsenal, AL 35898-5248		10. SPONSORING/MONITORING AGENCY REPORT NUMBER
11. SUPPLEMENTAL NOTES		
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public Release; SBIR report, distribution unlimited		12b. DISTRIBUTION CODE
13. ABSTRACT (Maximum 200 words) During the SBIR Phase II contract, OPTRA has explored to viability of planar encoder technology as the metrology system for use in next-generation microlithography machines. Planar encoders offer the particular advantage of very high repeatability and virtually complete immunity to the effects of atmospheric turbulence. In this report we present data and analyses that delineate the strengths and weakness of the planar encoder metrology, particularly in comparison to laser interferometer metrology. With the support of this program, OPTRA has developed its NanoGrid metrology system, which is now commercially available and being used in both the US and abroad. This system has sub-nanometer resolution, uncorrected accuracy of better than 100nm, and repeatability of better than 2.5nm. We describe methods of integrating the NanoGrid sensor with a state-of-the-art digital motion control system. Ultimately, the uncorrected accuracy that can be attained with this technology is limited by the accuracy with which the encoder grids can be manufactured. Many different techniques for grid fabrication have been explored and are described. Techniques for making large encoders, and for covering large areas with a small encoder and multiple sensor heads, were explored and are also described. Finally, a short-range non-contact Z-displacement sensor was developed and made into a commercial product; its design and performance are described.		
14. SUBJECT TERMS Encoder Metrology Planar Encoder Lithography		15. NUMBER OF PAGES 61 16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED		20. LIMITATION OF ABSTRACT UL

Introduction

This is the final report for DARPA SBIR program "High Speed, High Accuracy Stage for Advanced Lithography". The goal of this program was to explore the viability of planar encoder technology as the metrology system for next generation lithography (NGL) systems. Over the course of this program, OPTRA has developed a deep and thorough understanding of the advantages and limitations of planar encoder technology. In this report, we will document our understanding of planar encoder technology, present data that demonstrates its capability and present conclusions regarding its suitability for use in NGL systems. In addition, we will describe the NanoGage, a z-displacement sensor partially developed and commercialized under this program as a complement to the NanoGrid planar encoder.

This report is broken into four basic sections:

Section 1. Overview of Planar Encoder Technology In this section we describe in great detail the design and capability of the OPTRA NanoGrid sensor.

Section 2. Sensor Performance In this section we present data demonstrating the capability of the NanoGrid sensor and discuss some integration considerations..

Section 3. Applying Planar Encoder Technology to Next Generation Lithography In this section we compare the metrology needs of NGL with the demonstrated performance of the OPTRA NanoGrid system.

Section 4. The NanoGage In this section we describe the design and performance of the NanoGage sensor, a non-contact laser-based z-position sensor that was partially developed under this SBIR program.

Section 1. Overview of Planar Encoder Technology

1.1 Motivation, rationale, and performance goals

Until recently, the metrology systems used to accurately control XY stage position were either linear encoders on stacked stages, or laser interferometers used in conjunction with an L-shaped mirror (figures 1 and 2). In many applications requiring XY positioning, the goal is to position the stage accurately with respect to a point-of-action—this might be the location of a focused laser beam, the center of the field-of-view of a microscope, or the intersection of an electron or ion beam with the surface of the stage. Ideally, the X and Y measurements should be made as close as possible to the point-of-action—to minimize Abbe errors and to take thermal expansion of the stage out of the picture.

The motivation in developing a planar encoder (or XY encoder) metrology system was to achieve accuracy and repeatability at the 1 nm to 10 nm level in a system that was compact and economical. The difficulties in stacked stages with linear encoders are (1) the play in the bearings connecting the upper and lower stages creates a non-repeatable position error, and (2) the points of measurement are necessarily well-separated from the point-of-action. The difficulties with a laser interferometer metrology system are (1) unless the optical paths within the interferometer are in a vacuum, air turbulence (both ambient and that caused by the stage motion) produces non-repeatable position errors, and (2) laser interferometers tend not to be compact, and their points of measurement are displaced from the point-of-action.

The major commercial requirement for accurate stage positioning (at the nanometer level) has been microlithography in connection with semiconductor device fabrication, and accurate wafer positioning for semiconductor memory repair. Both of these applications require accurate positioning in combination with a large dynamic range; 10nm positioning accuracy and repeatability in combination with a 300mm range of motion is a typical high-end requirement.

In developing an XY encoder metrology system, our rationale was to use a laser interferometer in a highly controlled environment (i.e. negligible turbulence errors) to make a large XY encoder on a stable substrate, which could subsequently be used in an uncontrolled environment to provide repeatable and accurate position measurements. Our technique for reading the XY position of the encoder is essentially interferometric, but the optical paths are very short and very close to one another, so that the effects of air turbulence are negligible.

By using a laser diode for making encoder position measurements, we can achieve very high signal-to-noise ratios, and can therefore achieve the deep interpolations needed to work with relatively coarse encoders: we are currently using encoders with a 10-micron mechanical pitch, and a 5-micron optical pitch. This coarse pitch (coarse compared to wavelength of light, which defines the pitch for a laser interferometer) allows us to handle high stage speeds with relatively low signal frequencies.

1.2. Principles of operation

1.2.1 Two-beam interference

As we will see, our method for measuring encoder position is to generate a sinusoidal signal as the stage moves, and to infer the stage position from the phase of the sinusoidal signal. The sinusoidal signal goes through one complete cycle for each 5 microns of stage displacement. We wish to measure position with a resolution of 0.3nm, corresponding to an interpolation of 1 part in 16,384 or 14-bits. This means that the signal has to be very nearly a perfect sinusoid in order to allow accurately linear measurements. We generate this signal from a two-beam optical interference pattern, and are consistently able to make measurements with a residual (repeatable) interpolation error of less than 5nm. Figure 3 shows the basic layout of a single axis sensor head of the type we use. A beam from a laser diode is incident on the encoder, and the diffracted light is distributed between the various diffraction orders—which are separated in angle by $\Delta\theta \approx \lambda/d$, where λ is the wavelength of the laser diode and d is the encoder spacing. How the light is distributed between the various orders depends of the structure of each cycle of the encoder, but not on its period. As shown in figure 3, a spatial filter is used to select the ± 1 diffracted orders, and a lens is used to bring these two beams together where they interfere to form a purely sinusoidal interference fringe pattern. The spatial period, d' , of this fringe pattern is given by:

$$d' = \frac{M\lambda}{2\Delta\theta} = \frac{Md}{2}, (1)$$

where M is the magnification of the imaging lens.

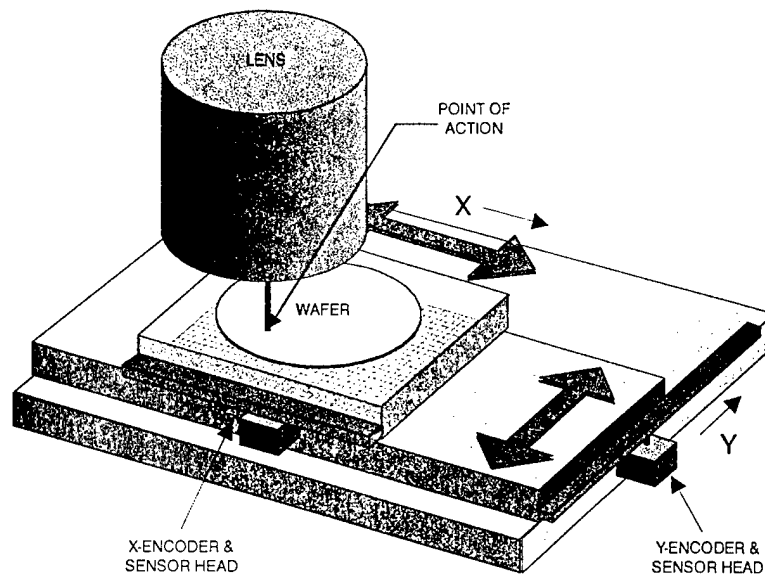


Figure 1 Stacked stages & Linear encoders

If linear encoders are used in an XY-metrology system, they must be used on stacked stages as shown here. The primary disadvantage of this arrangement is that there is an entire set of stage bearings between the lower encoder measurement and the part whose position is being measured. Since it is very difficult to make stacked stages repeatable to better than a few microns, the accuracy of such stages is similarly limited. Moreover, since the bearing errors are generally not repeatable, these errors cannot be calibrated.

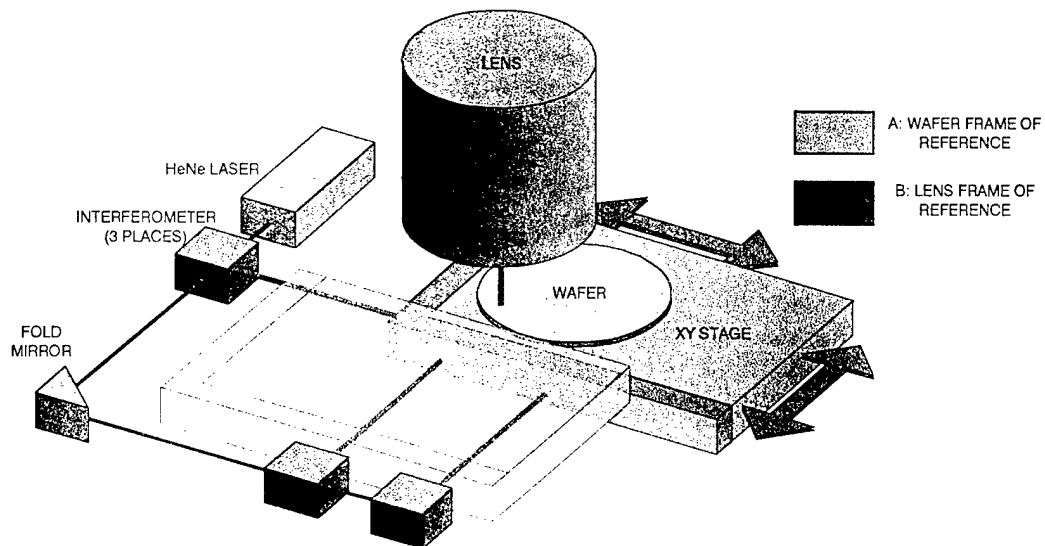


Figure 2 Laser Interferometer Metrology System

This figure illustrates a typical laser interferometer installation, with a laser, three interferometer heads, and an L-mirror on the XY stage. The range of motion of the stage is illustrated by the light phantom stage. The components shown in blue are in the stationary reference frame, and those shown in red are in the moving reference frame.

As can be seen, the laser interferometer is relatively easy to add to an XY stage system, but the footprint of the metrology system is large, and the air paths are at least as long as the maximum stage displacement.

1.2.2 Measuring the phase of the interference fringe pattern

The general expression for the intensity of light in an interference pattern is:

$$I = I_1 + I_2 \cos(2\pi x / d' + \phi_0), \quad (2)$$

where I_1 is the DC light level, I_2 is the amplitude of the spatial modulation, x is the position at the point of measurement, d' is the fringe spacing, and ϕ_0 is a constant phase. The values of I_1 and I_2 depend on the intensities of the two interfering beams. In an interferometer, these values depend on the reflectivity of the mirrors and may vary with position and the direction in which the light is reflected.

In AC interferometry the interfering beams are at slightly different frequencies and the interference pattern is described by:

$$I = I_1 + I_2 \cos(4\pi x / \lambda + \phi_0 + 2\pi \Delta f t), \quad (3)$$

where Δf is the laser difference frequency. As long as the mirror speed is less than $2\Delta f / \lambda$, the detector output can be AC-coupled so that, in effect, equation (3) becomes:

$$I = I_2 \cos(4\pi x / \lambda + \phi_0 + 2\pi \Delta f t), \quad (3a)$$

so that the variable I_1 is eliminated. However, in general, the I_2 term is slightly modulated at the beat frequency Δf (cross-talk) and there are associated measurement errors.

We have accepted the fact that there are really three variables in equation (2), namely I_1 , I_2 , and x , and in our sensor head we make three measurements:

$$R = I_1 + I_2 \cos(2\pi x / d' + \phi_0), \quad (4a)$$

$$S = I_1 + I_2 \cos(2\pi x / d' + \phi_0 + 2\pi / 3), \text{ and } (4b)$$

$$T = I_1 + I_2 \cos(2\pi x / d' + \phi_0 + 4\pi / 3). \quad (4c)$$

As shown in figure 4, these measurements are made with a specially fabricated tri-phase detector. This detector has 90 elements, each 9 microns wide on 15 micron centers. Every third detector element is connected in parallel, and the three detector outputs provide us with measurements of R , S and T in equation (4). These three signals are combined algebraically to give quadrature signals X and Y (not to be confused with the x and y values of the encoder position):

$$X = \frac{\sqrt{3}}{2}(S - T), \text{ and } Y = R - \frac{1}{2}(S + T). \quad (5)$$

Finally, the phase ϕ is found from:

$$\phi = \arctan(Y/X), \quad (6)$$

and the position x is given by:

$$x = \frac{\phi d}{4\pi}. \quad (7)$$

This is, of course, the value of x modulo 2π ; as the encoder moves we must keep track of the integral number of cycles, N , and report out a position x given by:

$$x = \frac{d}{2} \left[N + \frac{\phi}{2\pi} \right]. \quad (8)$$

(The assumption here is that the measurement was initialized at a point $x=0$ and that measurements are made relative to this point.)

This tri-phase measurement process has been highly effective in allowing accurate measurements even when the reflectivity of the encoder has varied over a range of 10:1. Moreover, because the measurement is averaged over 30 fringes, it is relatively insensitive to small dirt particles on the encoder, or local defects in the encoder pattern. The arctangent computation in equation (6) is carried out either by a 14-bit Pythagorean processor chip¹ or by means of an interpolated look-up table. We have also developed an ASIC that can measure ϕ with 8-bit accuracy using a resistor network and gate array.

1.3. Modeling performance

In making accurate position measurements, we are critically dependent on the precise measurement of the three signals R, S and T. In order to understand the role of all of the parameters involved (e.g. laser beam size and shape, overlap of the two beams on the detector, relative powers in the two interfering beams, position of the beams on the detector, size and structure of the detector, gains and offsets of the amplifiers in all three channels, etc.) we found it necessary to model the laser beams and detector in great detail. In this way we were able to determine how well matched the amplifiers had to be, and how the detector aperture should be apodized, in order to achieve a specified measurement linearity. In the model, the detector was divided into 270 equal-width strips covering the 90 detector elements; each strip could be assigned a separate length and sensitivity to light. The two interfering light beams were described by their center locations, their relative intensities, and by their Gaussian beam spot diameters. The 5 amplifiers (R, S, T, X, Y) were characterized by separate gains and offsets, and the magnification of the fringe pattern (relative to the detector spacing) could be changed by a variable factor. Figures 5, 6 and 7 show some of the results obtained from this model.

¹ The TMC2330AH5C manufactured by Raytheon.

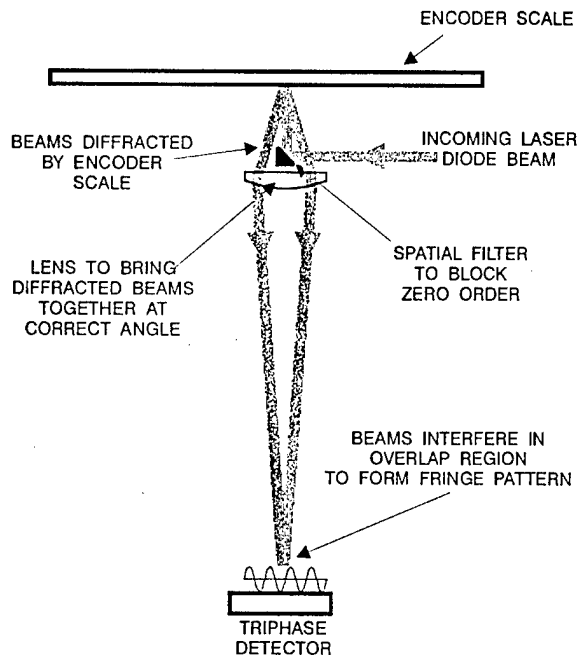


Figure 3 Interferometric linear encoder

The incident beam from a laser diode is diffracted at the encoder and the ± 1 diffracted orders are collected by the lens and brought together to form an interference pattern at the detector. (This interference pattern can be regarded as a spatially filtered image of the encoder; it has twice the normal spatial frequency because the zero-order has been removed by spatial filtering.) Because just two beams are interfering, the fringes have a perfectly sinusoidal intensity distribution. The fringe spacing at the detector is independent of the laser wavelength (except for the very small effect of chromatic aberration in the lens).

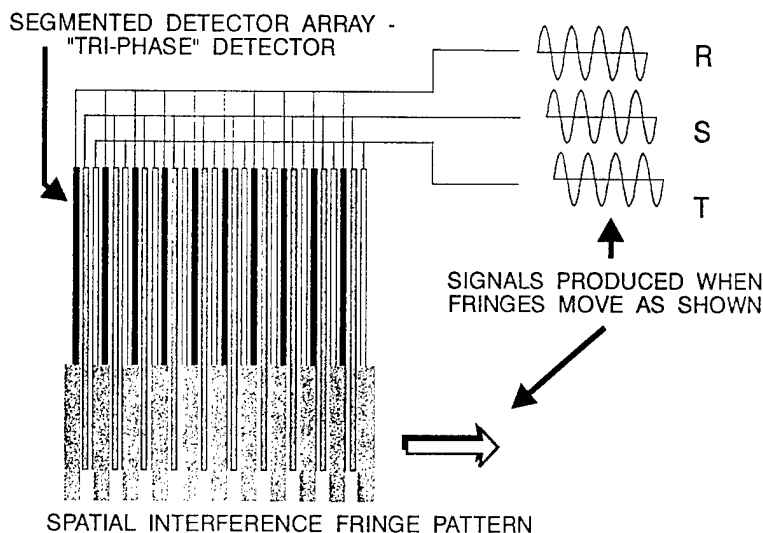


Figure 4 The tri-phase detector

This figure illustrates the tri-phase detector--so-called because when the encoder moves it produces three separate signals that are exactly 120° out of phase with one another. The detector consists of 3 sets of 30 elements each, arranged as shown, with every third element connected in parallel. The width of each element is 9 microns, and the separation between the centers of adjacent elements is 15 microns. The 3 separate measurements allow the phase of the fringe pattern (which has a 1:1 correspondence to the position of the encoder) to be measured independent of fluctuations in the laser power or variations in the relative amounts of power in the two interfering beams.

linencod.cdr

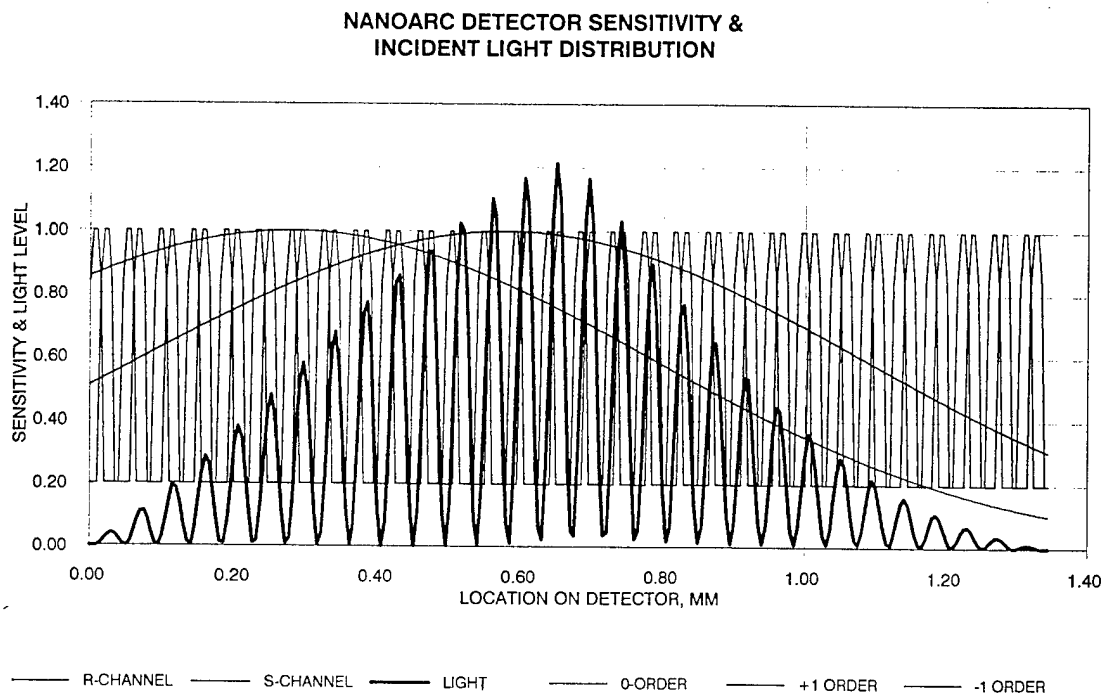
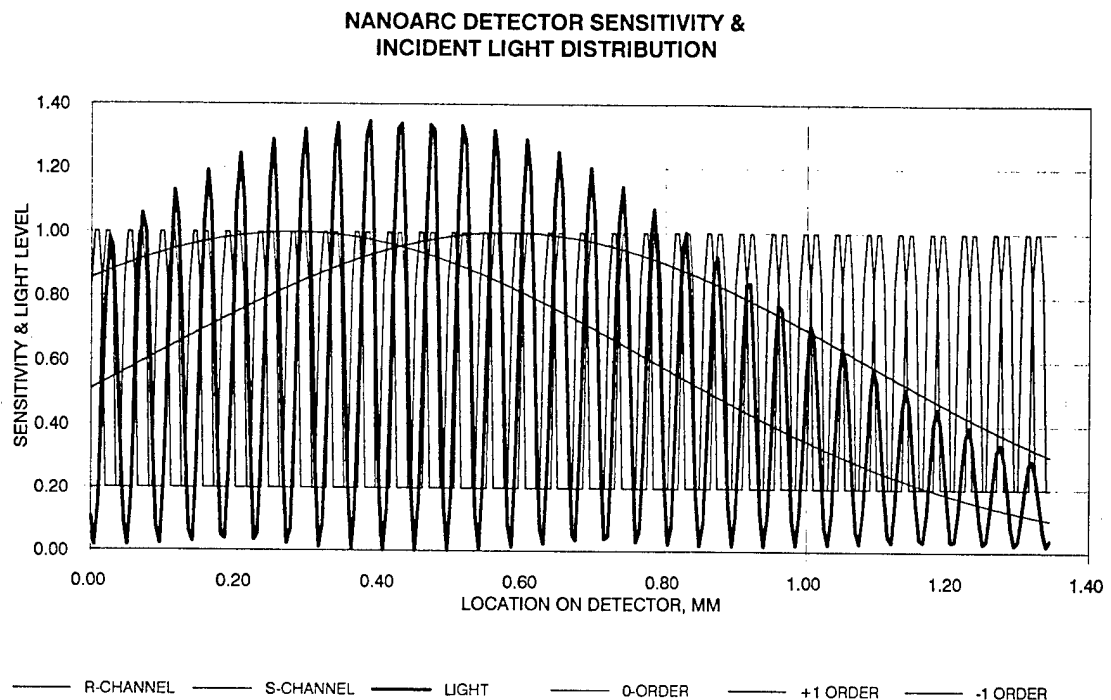


Figure 5 Modeled Light Distribution & Detector Sensitivity

The dark blue curves show the amount of light incident on the detector elements as a function of position on the detector. The upper plot models an un-apodized detector, while the lower plot represents a detector which has been apodized by a diamond-shaped mask (to smoothly reduce the amount of light at the left and right edges of the detector).

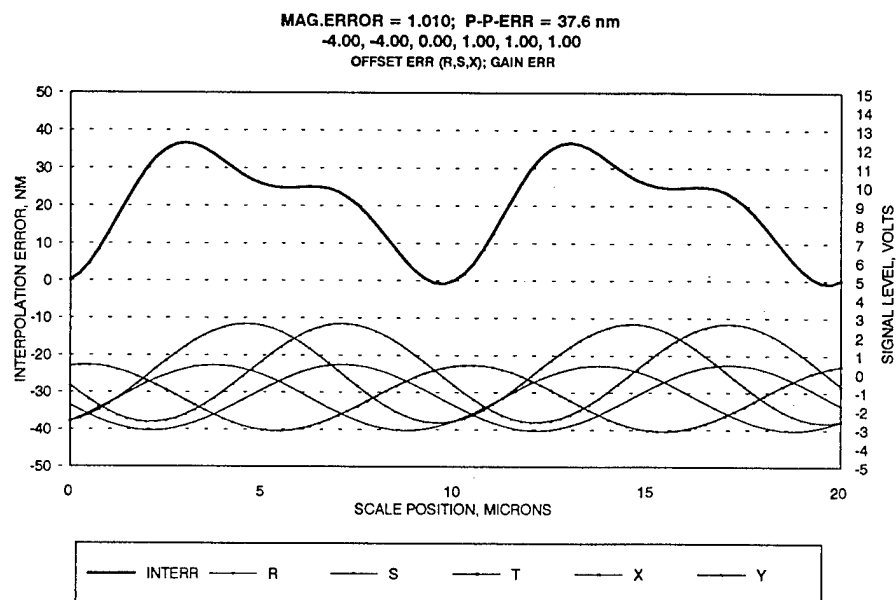


Figure 6 Modeled interpolation error, no apodization

Due to the asymmetry in the fringe pattern, combined with the high light levels at the edges of the detector, there is a significant periodic interpolation error (≈ 40 nm peak-to-peak). Experimental results were consistent with the model predictions.

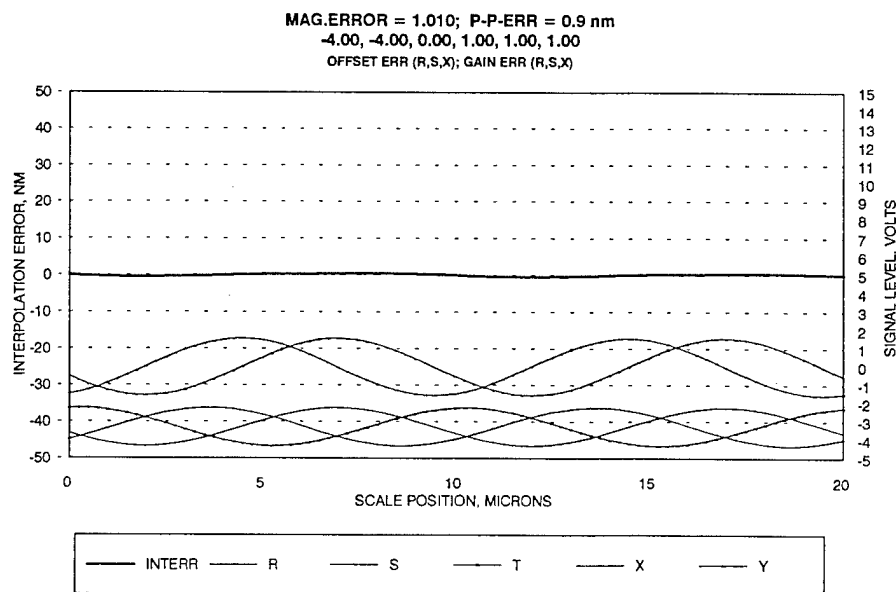


Figure 7 Modeled interpolation error; diamond aperture apodization

Apodization has reduced the interpolation error to less than 1 nm. In practice, we have been able to consistently keep the interpolation error well below 5 nm pk-to-pk. The plots shown in figures 6 and 7 above correspond to the light distributions shown in figure 5. The spreadsheet model upon which these data were based gave us both insight into the behavior of the NanoGrid sensor, and allowed us to address specific sensor problems.

1.4. Metrology error sources and their control

In this section we define each of the major metrology error sources and discuss the means for controlling them. The error sources that we will be concerned with are: encoder errors, Abbe error, interpolation error, velocity error, detector shot noise error, encoder obstruction error, and "other" errors.

1.4.1 Encoder errors

Encoder errors, i.e. errors in the positions of the rulings on the encoder, are the most fundamental errors that we have to contend with. Clearly, these are systematic errors and could, at least in principle, be measured and corrected for. These type of correction is quite feasible for errors with low spatial frequency, but becomes unmanageable at high spatial frequencies where the error table becomes very large and where initialization becomes critical.

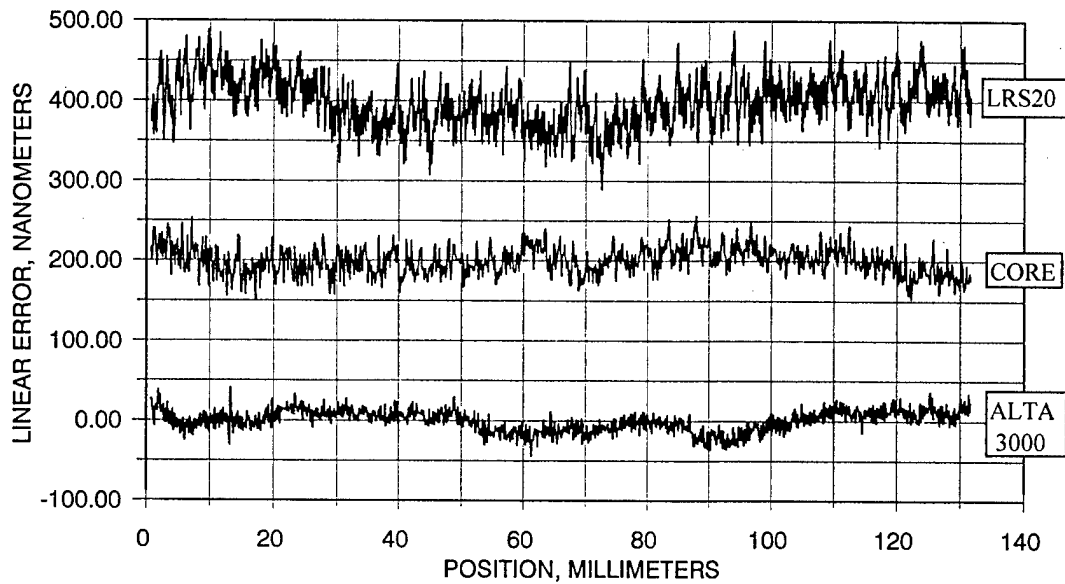
Our approach to the fabrication of encoders has been to make them as accurately as possible and to particularly avoid error sources with high spatial frequencies. Factors that invariably contribute to encoder errors are mechanical motions during fabrication, and the passage of time during the fabrication process. Mechanical motions lead to vibrations and small angular displacements that translate into Abbe errors, and the passage of time leads to varying environmental factors that affect encoder accuracy—such as temperature and atmospheric pressure. Following is a list of encoder fabrication techniques together with notes on their pros and cons in terms of accuracy.

FABRICATION TECHNIQUE	PROS	CONS
Step & repeat lithography	Standard practice; well-understood; machines are available to make large encoders ($\geq 400\text{mm}$ dimensions)	Inevitable stitching errors (because the metric of the projected image patches can never precisely match the metric of the stage motion)
Direct laser-write lithography	Lithography machines are available that provide excellent accuracy and absence of high-spatial-frequency errors (e.g. the CORE and ALTA machines made by Etec.).	The highest quality machines can only cover up to $175\text{mm} \times 175\text{mm}$; lower quality machines have quite large high-spatial-frequency errors. All machines have a scale factor error, typically a few parts in 10^6 .
Holography	Simultaneous and rapid exposure of the entire encoder go a long way to eliminate errors due to mechanical motions and long fabrication times.	Prohibitively expensive to make large encoders; requires two orthogonal exposures; requires large optics of very high quality. ²
E-beam lithography	Capable of very high spatial resolution, but this is not really a requirement for XY encoders with resolution at the nanometer level.	Not readily adaptable for the fabrication of large encoders.

We routinely measure encoder errors by comparing encoder measurements to those made with a laser interferometer, while taking care to minimize Abbe errors in either metrology system. Figure 8 shows error plots obtained with a variety of encoders. During these measurements, every effort is made to minimize atmospheric turbulence in the vicinity of the laser interferometer; the stage is moved slowly, and the stage and entire metrology system is enclosed in a plastic tent to reduce air currents and acoustic pick-up.

² We have developed a techniques for making long ($> 100\text{cm}$) holographic encoders without the need for large optics, but are not able to extend this technique to 2 dimensions. (To be published.) Please see appendix A for more details.

ALTA 3000 vs. CORE vs. LRS200 GRID CALIBRATION 6/16/98



IMPROVED STEP & REPEAT

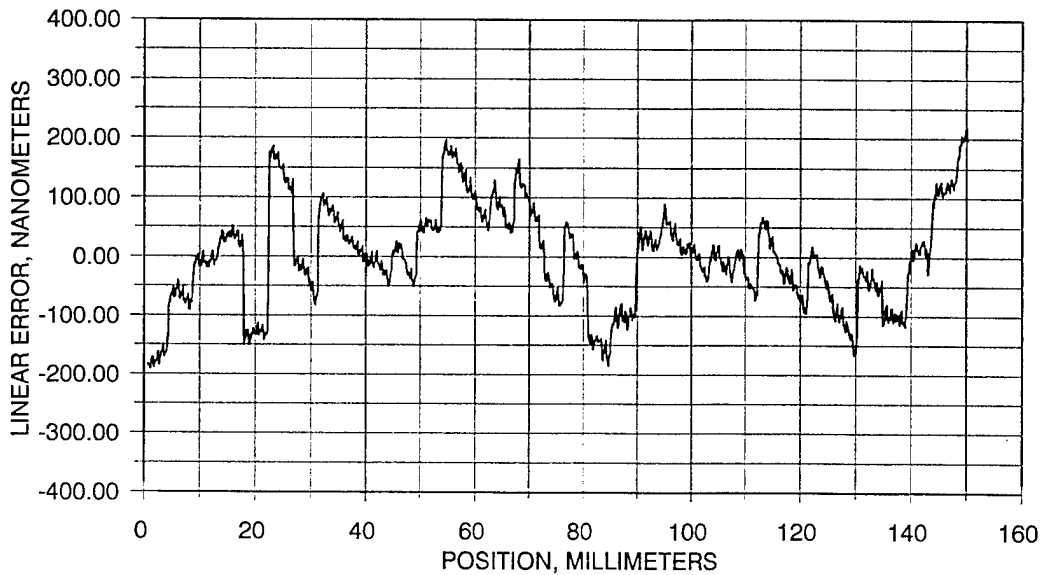


Figure 8. Error Plots for Various Encoders

The top plot shows error curves for grids made with three different direct-write techniques. The bottom plot is the best effort yet that we've seen for a step and repeat process. Clearly the ALTA 3000 offers the best overall accuracy. Unfortunately, it is limited to a plate size of 175mm. The step and repeat grid has the worst accuracy of the lot, but it is capable of handling up to 400mm plates.

Figure 9 shows measurements made with both an XY encoder and a laser interferometer on a nominally stationary stage in a normal laboratory environment (with no plastic tent). The effects of low-frequency air turbulence on the interferometer measurement is quite clear, while the encoder measurement makes it possible to see the small and slow displacement of the stage—probably due to thermal effects. While the laser interferometer is more accurate than the encoder (particularly in terms of systematic errors), the encoder measurements are more repeatable. Our ultimate goal is to capture the accuracy of a well-stabilized laser interferometer in a robust encoder on a fused silica substrate, and to thus combine the inherent accuracy of the laser with the inherent stability of the encoder.

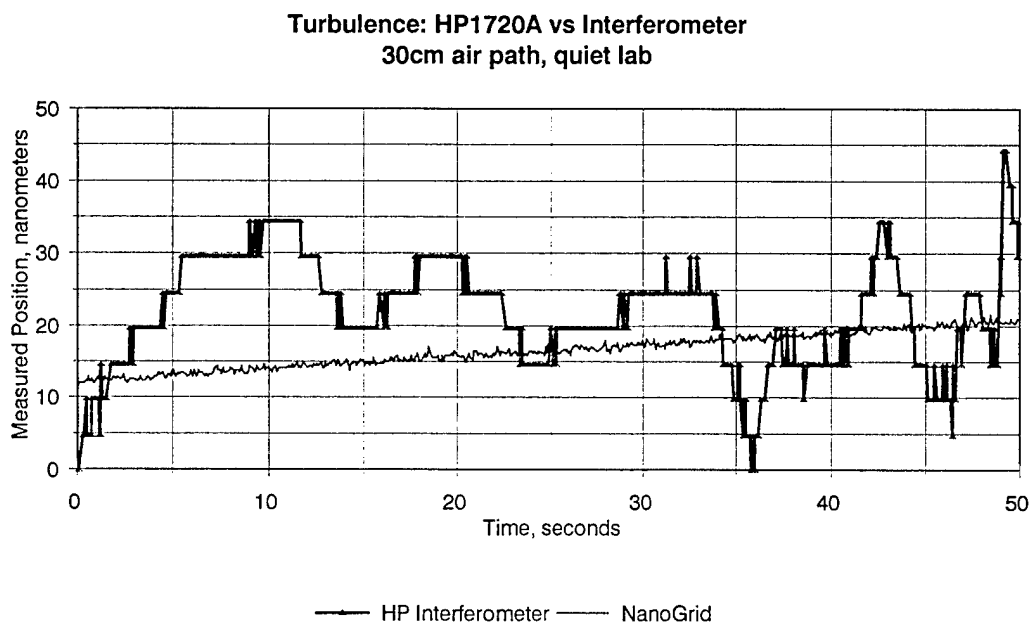


Figure 9 Stationary stage; effects of turbulence on position measurement

Normal quiet laboratory environment. Residual air turbulence causes a 40nm peak-to-peak random error in the laser interferometer measurement, while the NanoGrid measurement is essentially unaffected. Note how the slow displacement (probably thermal in origin) can be seen in the NanoGrid measurement, but not in that of the laser interferometer.

1.4.2 Abbe error

Abbe error occurs when (a) there is a physical separation between the point of measurement and the point of action (we define the *point of action* as the point at which we desire the measurement), and (b) there is a rotation of the moving frame of reference relative to the stationary frame of reference. For the present discussion, we will assume that the rotation is about the Z-axis, and that the desired measurements are of the X and Y displacements of the moving frame of reference (e.g. the frame containing the wafer, the XY encoder, and the stage that carries them) relative to a point of action located in the fixed frame of reference (e.g. the tip of an atomic force microscope). Figure 10 illustrates the situation. The point of action is at P, and the point of measurement is at P'. If the moving frame of reference rotates about the Z-axis by an angle $\Delta\theta$, as shown, there is no X or Y displacement of the point P, but there will be displacements Δx and Δy at P'.

$$\Delta x = y\Delta\theta \text{ and } \Delta y = x\Delta\theta, \quad (9)$$

where x and y are the coordinates of the point of measurement, P', relative to the point of action, P.

Note that if we are concerned only with displacement in the y-direction (vertical) then the Abbe error can be reduced to zero by making sure that P and P' lie on a line that is parallel to the y-axis.

Abbe errors are insidious because they can be either repeatable or non-repeatable. If they are repeatable, then they can easily be misinterpreted as encoder errors—and if they are non-repeatable, they can be misinterpreted as turbulence (or other non-systematic) errors. Repeatable Abbe errors are commonly caused by small imperfections in the ways of an air-bearing stage, while stages with circulating bearings generally cause non-repeatable Abbe errors. An advantage offered by an XY-encoder metrology system is the ability to locate the point of measurement very close to the point of action. The ultimate reduction of Abbe error would be a metrology system in which the wafer whose position was to be measured had an XY-encoder on its rear surface, being viewed through a transparent fused silica stage with the point of measurement directly below the point of action.

To get an idea of the magnitudes involved in Abbe errors, consider a 1 mm lateral (x) offset between the point of measurement and the point of action. A 0.1 milliradian rotation (i.e. 1mm in 10m) about the z-axis would then cause an Abbe error in a y-measurement of $1\text{mm} \times 0.0001\text{rad} = 0.1\mu\text{m}$, or 100nm. As we shall see later, actual measurement errors of this magnitude are frequently due to Abbe error.

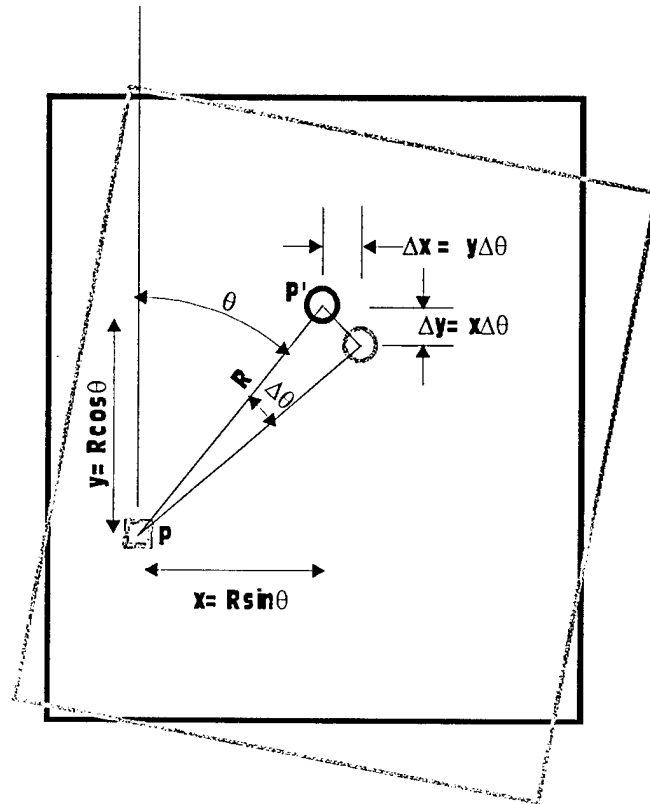


Figure 10 Abbe Error

P is the point of action (where the measurement is desired) and P' is the point of measurement. The black and gray portions of the figure show the stage before and after the stage has been rotated about the Z-axis at P by $\Delta\theta$ (rotation about another point would be equivalent to rotation about P plus a common Cartesian displacement). The changes in the measurement at P' due to the rotation about P is the Abbe error. Clearly, the Abbe error is minimized by keeping the point of measurement as close as possible to the point of action.

1.4.3 Interpolation error

As we saw in equations 4, 5, 6, and 7, the measurement of position (modulo d) is inferred from measurements of R, S and T—predicated on the assumption that the values of I_1 and I_2 in equations 4a, 4b, and 4c were all equal. In fact, due to mismatches in the gains and offsets of the detectors and amplifiers that create these signals, this assumption is never fulfilled. We have termed the resulting errors *interpolation errors*, and have found that by keeping the gains and offsets them matched to within 0.1%

(which we can just do using standard electronics components), we are able to keep the interpolation error below 10nm peak-to-peak.

The form of the interpolation error depends on its origin. Gain mismatches between R, S, and T signals cause a sinusoidal interpolation error with a period of $d/4$, and offset mismatches produce an interpolation error with a period of $d/2$. If there is leakage of zero-order diffracted light onto the detector, there tends to be an interpolation error with a period of d . Errors in the gains and/or offsets of the quadrature phase signals X and Y cause similar interpolation errors.

1.4.4 Velocity error

Velocity error is a direct result of the finite *data lag*, or time lag between the occurrence of a position value and the corresponding output signal. If the stage is moving in the x-direction with a velocity v_x , and if there is a data lag τ , then when the stage position is reported, the position measurement of x will be in error by:

$$\Delta x_{\text{VELOCITY}} = v_x \tau. \quad (10)$$

For a 10 μ sec value of τ and a stage speed of 100mm/sec, the velocity error is 1 μ m.

Fortunately, the velocity error is easy to correct for and this is routinely done. If, however, the data lag is variable, then the velocity cannot be corrected for and there may be significant errors if measurements are made when the stage is moving in different directions or at different velocities.

1.4.5 Shot noise error

For a given nominally constant optical power falling on a detector, there is a corresponding *shot noise* associated with the randomness in the number of photons detected within any specified time interval. When, on the average, N photons are detected within a specific time interval, there is an rms variability $\delta N_{\text{RMS}} = \sqrt{N}$, in the number of photons detected in a single measurement. This is a fundamental noise source, and the associated noise can only be reduced by increasing the light level or the measurement duration (bandwidth reduction). A well-designed sensor system is generally *shot-noise limited*, meaning that all of the controllable noise sources have been reduced to a level below that of the shot noise.

4.6 Grid-obstruction error

Large defects or debris on the grid surface can cause a measurement error. In the worst case, the debris can completely occlude the output laser spot, thus causing the sensor to lose count. This is referred to as a *global error* in that all subsequent measurements will be off by the number of counts missed during occlusion. In a less severe mode, debris can occlude a small portion of the illuminating spot, thus causing the phase of the fringe pattern to be shifted. These are known as *local errors* in that subsequent measurements return to the non-occluded value.

For lithography applications, this error is dealt with by ensuring that the grid is made with no defects and that it is clean upon installation. The lithography process itself requires ultra-clean operation, so the risk of post-installation contamination is small. For other applications, however, this may not be the case. For instance, in memory repair, debris is generated as part of the repair process. While every effort is made to keep this debris from the vicinity of the grid, there is still the risk that the grid can become contaminated. The sensor design must take this into account.

The OPTRA sensor has been redesigned to minimize the effects of contamination. The primary design change has been to increase the number of fringes being sensed and thereby average out the contamination induced errors. In the vicinity of contamination, the fringe pattern undergoes a phase shift that is not proportional to stage position, thus creating a measurement error. By sensing fringes outside of the vicinity of the contamination, we are able to average out the contamination induced phase error.

1.5. Detailed design of an encoder-based metrology system

1.5.1 Sensor head

The encoders that we use have a mechanical pitch of 10 microns and an optical pitch of 5 microns; that is to say, the physical spacing between rulings on the encoder is 10 microns, but that the interferometric fringes go through one full cycle for each 5 micron displacement of the encoder. Each of the 90 detector elements has a width of 9 microns and is separated from the adjacent element by a 6-micron space. Each triad of detector elements is thus separated from its neighboring triad by $3 \times (9+6) = 45$ microns. Since there are 90 individual detector elements, the detector spans 30 interference fringes.

A basic requirement of the sensor head is thus to illuminate the encoder with a beam that is nominally 150 microns in diameter, and to then image this 150 micron spot onto the detector with a magnification of $9\times$, so that each 5 micron optical period is projected onto a 45-micron wide triad of detector elements. This was illustrated in figure 3. A further design requirement was that there be a physical standoff, between the sensor head and the encoder, of at least 12mm—to allow the sensor head to view the encoder through a window if necessary and to allow a generous separation (as well as a generous tolerance in establishing that separation) between the stage (carrying the encoder) and the sensor head. The combination of the $9\times$ magnification and the 12mm minimum standoff has led us to the use of a singlet lens with a object distance of 18mm and an image distance of $9 \times 18\text{mm}$ or 162mm. The magnitude of these distances, particularly the image distance of more than 160mm, led us to a design with a number of fold mirrors in order to contain the optical path within a reasonable space. Accurate alignment of all of the components in the sensor head is critical, and we have had to develop an assembly procedure which, although well-automated, still requires skill and training on the part of the technician. Assembly of the sensor head is accomplished relative to reference surfaces within the sensor head, and these same reference surfaces are used to mount the sensor head relative to the encoder so that good overall alignment can be achieved entirely through fixturing—with no alignment adjustments required during installation in the field.

The sensor head contains not only the mirrors, prisms, and lenses needed to accomplish the optical design, but also two tri-phase detectors and associated amplifiers, plus the 785nm laser diode and its drive electronics, including electronics for AGC and potentiometers to balance the R, S, and T gains and offsets. A single cable connects the sensor head to either a PC plug-in ISA-card for high resolution (0.3nm) or a small electronics module for lower resolution (20nm) and the option for an A-quadrant-B output.

1.5.2 Detector

The detector is essentially 90 separate silicon PIN diodes, with every third diode connected in parallel and with three output signals (R, S, and T). Each single element is 1mm long and 9 microns wide, with a gap of 6 microns between adjacent elements. The

detectors work in photovoltaic mode with no reverse bias (in order to avoid a spurious dark current whose value gradually increased over time). Each detector is packaged in a TO-15 can, sealed with an AR-coated glass window, with 4 active leads (R, S, T and ground).

Surrounding the active region of the detector is an opaque black mask: opaque to keep light from reaching photosensitive regions other than the 90 detector elements, and black to reduce the amount of light reflected back from either the window or other more distant optical elements. We eventually found that if the image of the laser spot on the encoder fell onto the left or right edge of the detector (assuming the individual elements ran up and down), there would be an unacceptably large interpolation error (caused by in-balance in the R, S and T signal levels). Working with the spreadsheet model, we found that this problem could be very nearly eliminated by apodizing the detector aperture with a diamond-shaped mask. We subsequently found that reducing the sensitive area of the detector in this way had the undesired effect of increasing the sensor's sensitivity to dirt on the encoder. The choice of whether or not to use an apodized detector depends on which, in a particular application, was the greater problem: alignment stability (loss of which might steer the laser spot onto the edge of the detector), or dirt on the encoder.

1.5.3 The encoder, or grid

The ideal encoder has the following properties:

1. Uniform spacing of the diffracting features on the surface of the encoder;
2. Orthogonality of the features defining the X and Y axes;
3. Flat surface, so that the standoff doesn't vary with position and the incident beam is always perpendicular to the encoder surface;
4. High diffraction efficiency into the ± 1 orders for both X and Y;
5. Thermal and mechanical stability; and
6. Low or moderate cost.

We will deal with each of these issues in the following paragraphs.

1.5.3.1 Uniform spacing of diffracting features

This issue, of course, addresses accuracy and has been discussed at some length in section 4.1. For applications where the key requirement is repeatability (such as overlaying device features on different layers while the wafer remains on the same machine), accuracy may be a secondary issue. But in general, accuracy is the key requirement in a metrology system.

As discussed earlier, we have found that by making master encoders on an ALTA laser-writing microlithography system we have obtained our best results on grids up to 175mm \times 175mm. With care, master gratings can be replicated by pseudo-contact printing (the plates are not actually in physical contact, but are very close)—but this requires that both the master and the copy be very flat. In practice, replicated gratings are almost as expensive as masters and we have rarely chosen to use such replicas.

For larger XY encoders, up to 400mm × 400mm, we have used a step-and-repeat process in which a 30mm square pattern is projected onto a 5mm square on the encoder substrate at $[1/6] \times$ magnification. Between successive exposures, the encoder is moved by a distance of 2.5mm, so that every point on the encoder has two superposed exposures. This results in a 2× reduction in the magnitude of the stitching errors (but there are twice as many of them).

1.5.3.2 Orthogonality

Orthogonality is taken care of by the same laser interferometer metrology that we rely upon to provide accurate encoder spacing. We can quickly check orthogonality with a pair of encoders by looking at the moiré patterns between them when they are placed in contact. If a clean moiré can be obtained both initially and after one of the encoders has been rotated by 90°, then the orthogonality is OK. These moiré patterns can very quickly reveal a wealth of information about poor encoders, but reveal little about the quality of encoders whose errors are less than a few hundred nanometers.

1.5.3.3 Flat surface

Since the ALTA machines will not accept substrates that aren't flat to better than a few waves, surface flatness is not an issue with encoders made on these machines. More generally, however, lack of surface flatness can cause a number of problems. Chief among these is steering of the reflected and diffracted beams when the surface is not perpendicular to the incident laser beam. The root problem here is spherical aberration in the imaging lens. When the incident beam is normal to the encoder surface, the ± 1 diffracted orders are symmetrically disposed on either side of the optic axis and both beams arrive at the image plane without any difference in optical path. However, if the incident beam is not normal to the encoder surface, the reflected ± 1 diffracted orders pass through different zones of the lens and, if the lens has any spherical aberration, there will be an optical path difference and a corresponding shift in the position of the interference fringes on the detector. One optical wavelength of optical path difference will cause the interference fringes to move by one full fringe period—producing a measurement error of 5 μm . An optical path difference of only $1/100$ -wavelength will cause a measurement error of 50nm. For this reason we have taken precautions to minimize spherical aberration, and to keep the incident beam normal to the encoder surface (either of these measures could in principle eliminate measurement errors due to spherical aberration). The lens used to image the spatially filtered diffracted light onto the detector is an aspheric lens designed for zero spherical aberration. In our measurements with this lens we have been unable to detect any spherical aberration.

1.5.3.4 High diffraction efficiency into the ± 1 orders

Ideally, all of the diffracted and reflected light from the encoder would go into the X and Y ± 1 orders. This can, in fact, be very nearly achieved by using a reflective phase encoder in which half of the light reflected into the 0-order is just out of phase with the other half. This requires a grating profile in which half of the surface area is $1/4$ -wavelength higher than the other half. This condition ensures that the 0-order diffracted light will be eliminated. Further, by making the surface profile symmetrical, the even diffracted orders can be reduced to zero. Figure 11 shows such a grating profile, and figure 12 shows the intensity distribution in the resulting diffraction pattern.

To date we have not found it necessary to achieve the very high radiometric efficiency that is possible with the type of grating described above, and we have used an amplitude grating of the type shown in figure 13, having the intensity distribution shown in figure 14. Our reason for using this relatively inefficient encoder lies in its lower cost: by using such an encoder we can avoid the cost of custom coating of the encoder substrates with chromium having a post-processing thickness of exactly $\frac{1}{4}$ -wavelength, and the additional cost of adding a reflective coating on top of the processed encoder plate.

Finally, during this program we have investigated means of calibrating encoders in situ. As of this writing, the amount of data handling required to adequately smooth out the high frequency errors that we find in our grids is unmanageable. Please see appendix B for more details.

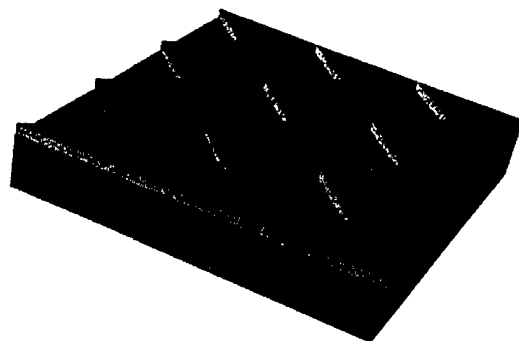


Figure 11. Phase Grid Profile

A grid fabricated with this pattern will act as a phase grating, completely suppressing the 0-order reflection from the grid and maximizing the ± 1 -order diffracted light.

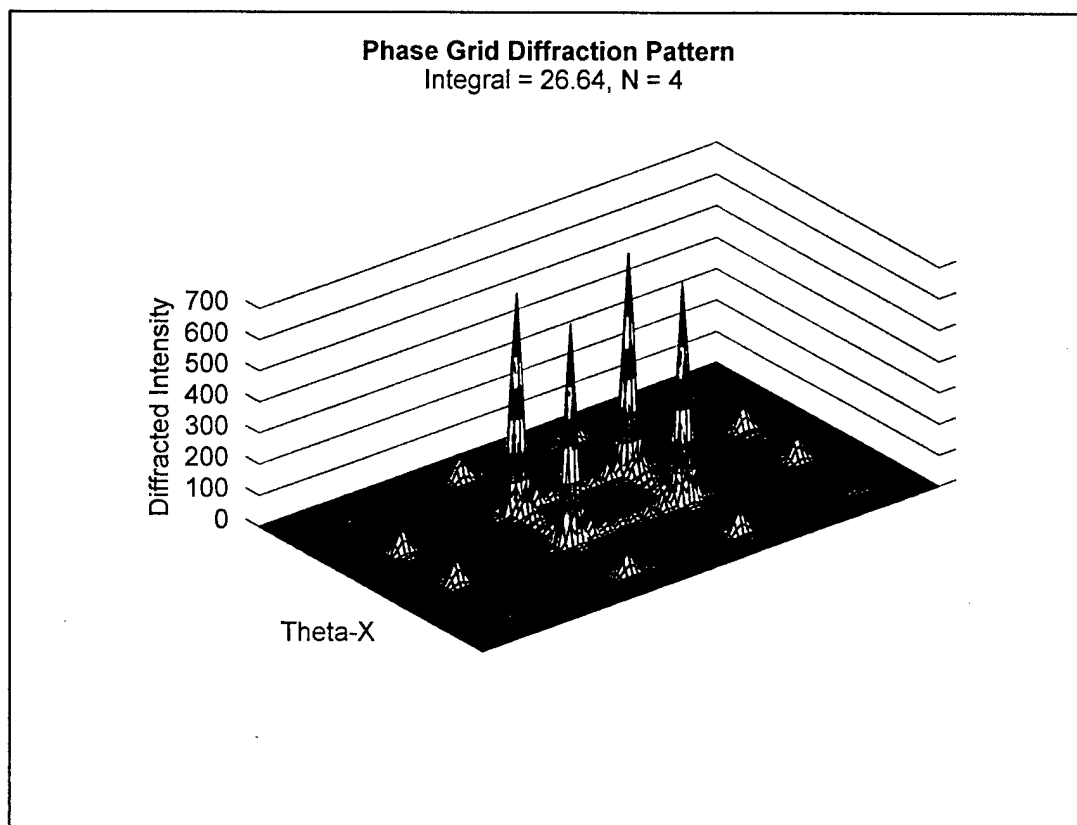


Figure 12. Diffraction Pattern from a Phase Grid

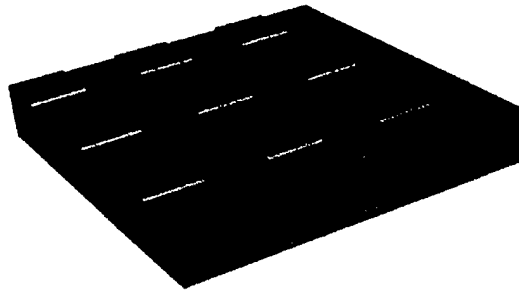


Figure 13. Standard Grid Profile

This type of grid allows a very strong 0-order reflection. It has the advantage, however, of being much less expensive to manufacture than the phase grid shown in figure 11.

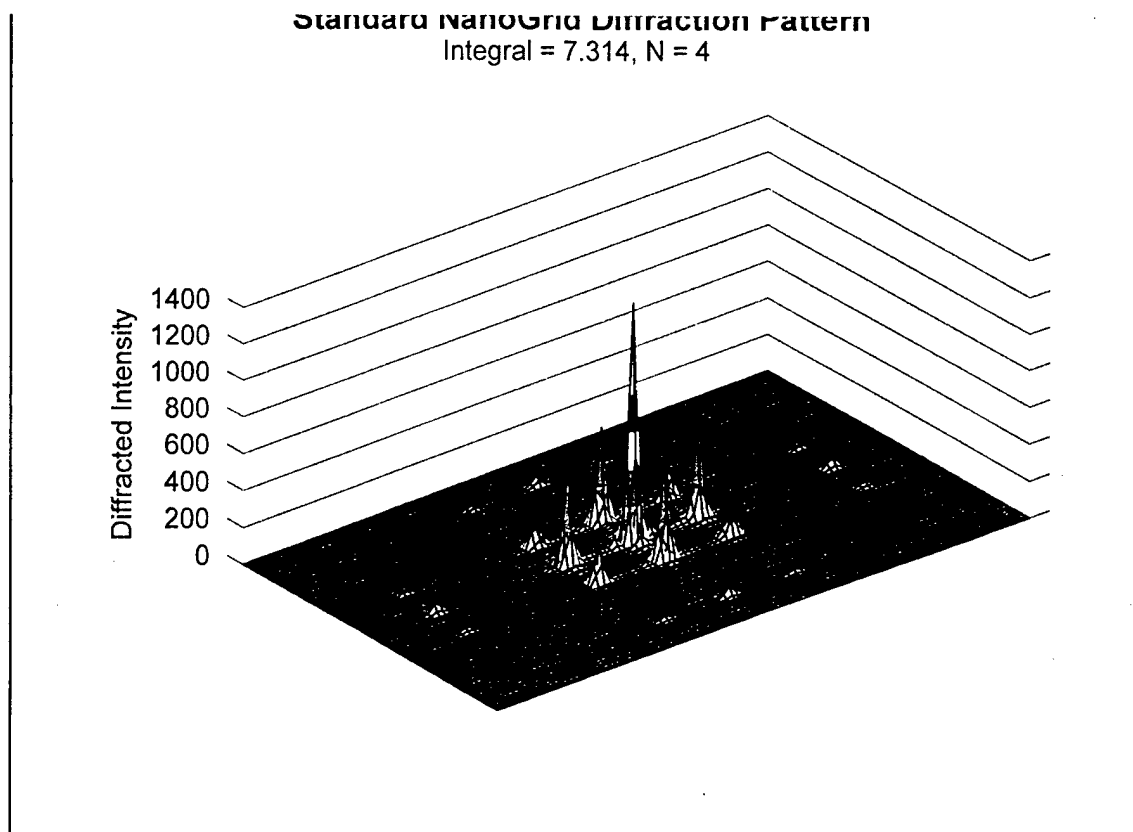


Figure 14. Diffraction Pattern from a Standard Grid

1.5.3.5 Thermal & mechanical stability

Thermal and mechanical stability of the encoder plate is clearly a requisite for accurate metrology. Thermal stability is achieved by the use of fused silica as a substrate material (coefficient of expansion $\pm 0.51 \times 10^{-6} \text{cm/cm}^\circ\text{C}$), and mechanical stability is achieved by the choice of substrate thickness and the means by which the encoder is attached to the moving stage. We generally mount larger encoders to an aluminum mounting plate by epoxying the encoder to three raised pads on the aluminum plate, each of which has a machined flexure which prevents distortion of the encoder when the mounting plate undergoes thermal expansion.

In some cases, for example where constant temperature cannot be maintained, it may be desirable to make the encoder of the same material as the target object whose accurate position is desired, and to place the target object directly on the encoder substrate. In this way there is no relative motion between the encoder and the target object when the temperature changes, and position measurements are automatically corrected for the temperature at which the encoder was calibrated (i.e. the temperature at which its linear correction factor was determined).

Generally, however, most applications requiring accuracy at the nanometer level provide a high degree of temperature control, so that thermal expansion is not a critical issue. When the temperature is not tightly controlled, all of the structural members of the stage and metrology system—particularly those which mechanically couple the sensor head to the stationary frame of reference, and which couple the encoder and target object to the moving frame of reference—are potential sources of temperature induced error.

1.5.3.6 Low or moderate cost

Our two primary goals in developing an XY metrology system that would provide an alternative to 2-frequency laser interferometer metrology systems were (1) to provide improved precision (i.e. repeatability) relative to a laser interferometer in the presence of atmospheric turbulence, and (2) to provide a lower cost system that could meet the needs of the growing market for metrology at the sub-micron level—where laser interferometers have in the past been the only available choice. The critical components of the laser interferometer metrology system are the wavelength-stabilized 2-frequency HeNe laser, the accurate L-mirror (accurately flat faces and an accurate 90° angle), and the interferometers. The critical components of the XY encoder based metrology system are the encoder and the sensor head—with the greater emphasis on the encoder. Note that for the XY encoder there is no requirement for laser wavelength stability; in fact, the system continues to function accurately even when the laser diode is operated below threshold as an LED.

A master XY encoder is expensive to make, but it has the clear potential for replication at quite modest cost. The highest cost item in a laser metrology system is the L-mirror, and this is a recurring cost; the highest cost item in the XY encoder metrology system is the master encoder, which is a non-recurring cost.

1.5.4 Electronics and software

The NanoGrid system has two basic electronic components: a preamplifier/laser drive assembly that resides in the sensor head and a 32-bit phase processor that is a stand-alone PC plug-in card.

1.5.4.1 The Preamplifier Assembly

This assembly amplifies the R, S and T signals, drives the laser diode, and prevents saturation via an AGC circuit. This assembly also provides a means of tuning the R, S and T signals to allow us to achieve interpolation errors of less than 10nm. During this program, the preamplifier assembly was modified to allow for bandwidth matching of the R, S and T channels. This was necessary to allow us to maintain low interpolation errors at high speed.

The new design has two stages. The first stage is a transimpedance amplifier with variable capacitors for matching bandwidths. Gain is determined by 0.1% resistors. The second stage is a non inverting amplifier with gain and offset adjustments. The circuits are packaged on two circuit boards, stacked on top of each other. The bottom board includes the transimpedance amplifiers. The two boards are connected by a 12 pin header at one end. This design has been implemented and tested successfully (see section 2.5)

1.5.4.2 The Hi-Res Phase Processor

The high-res processor has been modified to allow for operation at >1 meter/sec and a provision has been made for an external clock that allows complete synchronization of all data to a user-defined master clock. Two timing signals are provided by the user electronics to the interpolator card, these signals are a 20MHz clock and a 20MHz / 1120 = 17.857...kHz clock. The 20MHz clock is provided to JP4 pin 50. The 17.857kHz clock is derived from the 20MHz clock and is provided to pin JP3 pin 40. All data conversion performed on the interpolator is done synchronously to these 20MHz and 17.857kHz clocks.

Data conversion consists of counting interpolation pitches and interpolating within a pitch to produce a 32-bit position word. Note that the grid pitch of 10 μ M results in an interpolation pitch of 5 μ M. The lower 14-bits of the 32-bit position word are the "phase" within the current interpolation pitch and the upper 18-bits are the cumulative interpolation pitch count since the board was reset. The sensor range is $(218-1) * 5 \mu$ M pitch = 1.3 Meters.

X data is ready at the rising edge of XCLOCK, Y data is ready at the rising edge of the YCLOCK. The YCLOCK is the XCLOCK inverted. The XCLOCK and YCLOCK are derived from the 20MHz clock and are 56 of the 20MHz clock cycles (2.8μ s, 357.1...kHz). Position data is sampled and converted every XCLOCK and YCLOCK cycle. Data is sampled 86 of the 20MHz clock cycles before the rising edge of the corresponding XCLOCK or YCLOCK. The XCLOCK and YCLOCK are synchronous to the user supplied 17.857kHz clock. The rising edge of the XCLOCK occurs 6 of the 20MHz clocks (300ns) after the rising edge of the 17.857kHz clock. Note, X position data and Y position data are both sampled 20 times ($1120 / 56 = 20$) within one cycle of the 17.857kHz clock. (See Timing Diagram.) The interpolator provides signals on the JP3 and JP4 connectors.

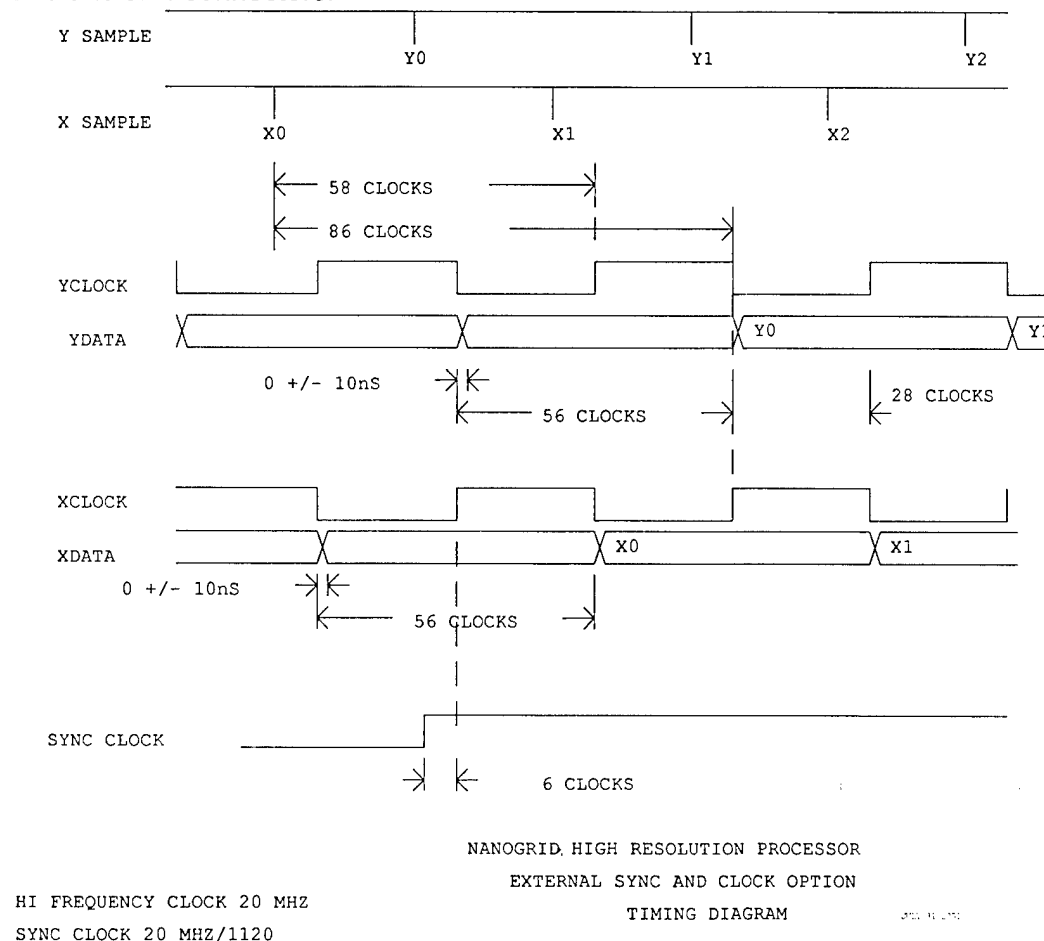


Figure 15. Encoder Timing Diagram

Criteria	Value	Comment
Ext Clk	20 MHz +/- 2 MHz	Interpolator will function correctly over

		a range of clock frequency.
Data Jitter	± 10 nsec (ref)	See Encoder Timing Diagram

2. Sensor Performance

6.1 Grid calibration & accuracy measurement

We have been able to calibrate and- measure the accuracy of XY encoders in one direction (X or Y) at a time. We have an air-bearing stage that can be servo-controlled in X and locked in Y. There is an overhead encoder sensor head to read X and Y, and there is a plane mirror laser interferometer to provide an independent readout of the X-coordinate of the stage position. The encoder is placed on the stage and its position is monitored by an overhead sensor head. The encoder is oriented so that when the stage is moved back and forth in X, there is minimum variation of the Y coordinate of the stage position as read by the encoder sensor head: this ensures that the encoder axes are parallel to the stage axes. In order to gather calibration (accuracy) data, the stage is slowly moved at constant velocity and both laser interferometer and encoder position measurements are made simultaneously at a specified frequency. Every effort is made to minimize Abbe error.

In a plot of encoder position vs. interferometer position, a linear fit to the data is made, and the encoder error at any point is taken as the difference between the actual encoder data and the linear fit to that data. This error is then plotted as a function of position in order to characterize the accuracy of the encoder for that particular fixed Y position. Data is obtained at a series of Y values, and the encoder is then rotated by 90° and the process repeated. The stage is moved slowly and is enclosed in order to minimize air turbulence during the measurement. Occasional data runs are repeated in order to check on the repeatability of the measurements. The data is generally repeatable to within 5 to 10 nanometers and we attribute this less than perfect repeatability primarily to residual air turbulence. The repeatable encoder errors are due in part to the actual errors in the encoder, and in part to small Abbe errors due to very repeatable steering of the stage as it is moved due to microscopic imperfections in the ways of the stage. (This issue will be addressed in more detail in a subsequent paragraph).

The slope of the linear fit to the encoder vs laser data is taken as a constant linearity factor and, in the case of encoders made with the ATLA machine, attests to a difference between the laser wavelength used in fabricating the grid, and the wavelength of the laser in order laser interferometer.

In fact, the metrology system used in making the encoders (combined with the high degree of averaging used in the lithographic process) is probably more accurate than our own metrology system. Nevertheless, our metrology system has enabled us to intelligently evaluate different encoder fabrication processes and to put an upper limit on the linearity of the encoders that we have used. We have also found that by taking the Fourier transform (in spatial frequency space) of our error data, we can identify the presence of mechanical vibrations during the lithography process.

2.2 Repeatability

In measuring repeatability we chose to make two independent encoder position measurements and to then assess the repeatability of the difference between repeated independent measurements. This approach let us completely avoid the non-repeatability

of the laser interferometer measurements associated with atmospheric turbulence. It also gave us the opportunity to assess the effects of data lag combined with velocity on measurement accuracy.

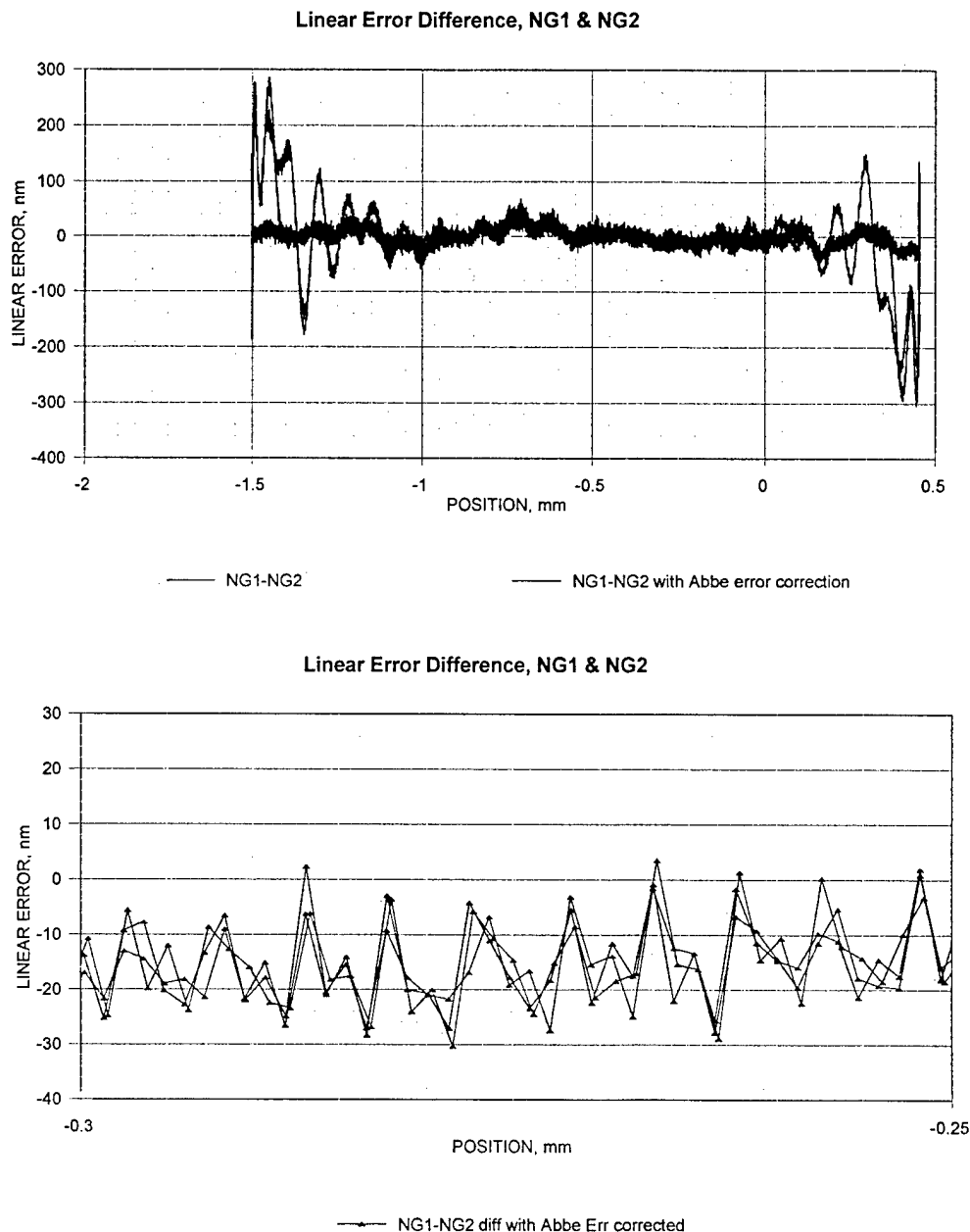


Figure 16 Repeatability of NanoGrid Measurements

As the stage was moved back and forth in x, its position was simultaneously recorded by two separate NanoGrid sensor heads. The metric for measurement repeatability was the extent to which the difference between the two sets of position readings was repeatable as the stage made a number of repeated movements. In the upper plot, the black curve is the raw measurement difference,

and the red plot has been corrected for Abbe error as monitored by yaw measurements in the y-direction. The lower plot is an expanded portion of the upper plot and shows a measurement repeatability on the order of 2nm.

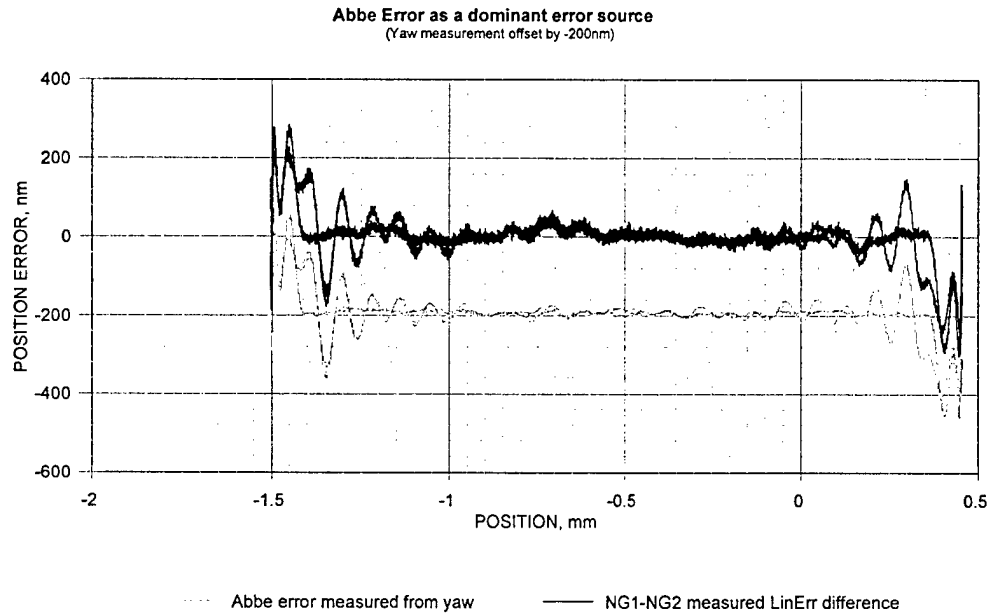


Figure 17 Abbe error as a major contributor to measurement error

In this plot the purple curve is a plot of the predicted Abbe error based on active measurements of the stage yaw (using the two y-axis sensors as the stage was moved nominally in the x-direction) and knowledge of the x and y components of the separation between the two NanoGrid sensor heads. The red plot shows the measured difference between the x-position measurements made with the two sensor heads. The two individual curves do not overlay because the yaw errors depended on which way the stage was moving—in other words, the yaw errors were repeatable, but only when the direction of motion was taken into account.

Clearly, the Abbe error associated with stage yaw is a dominant source of measurement error.

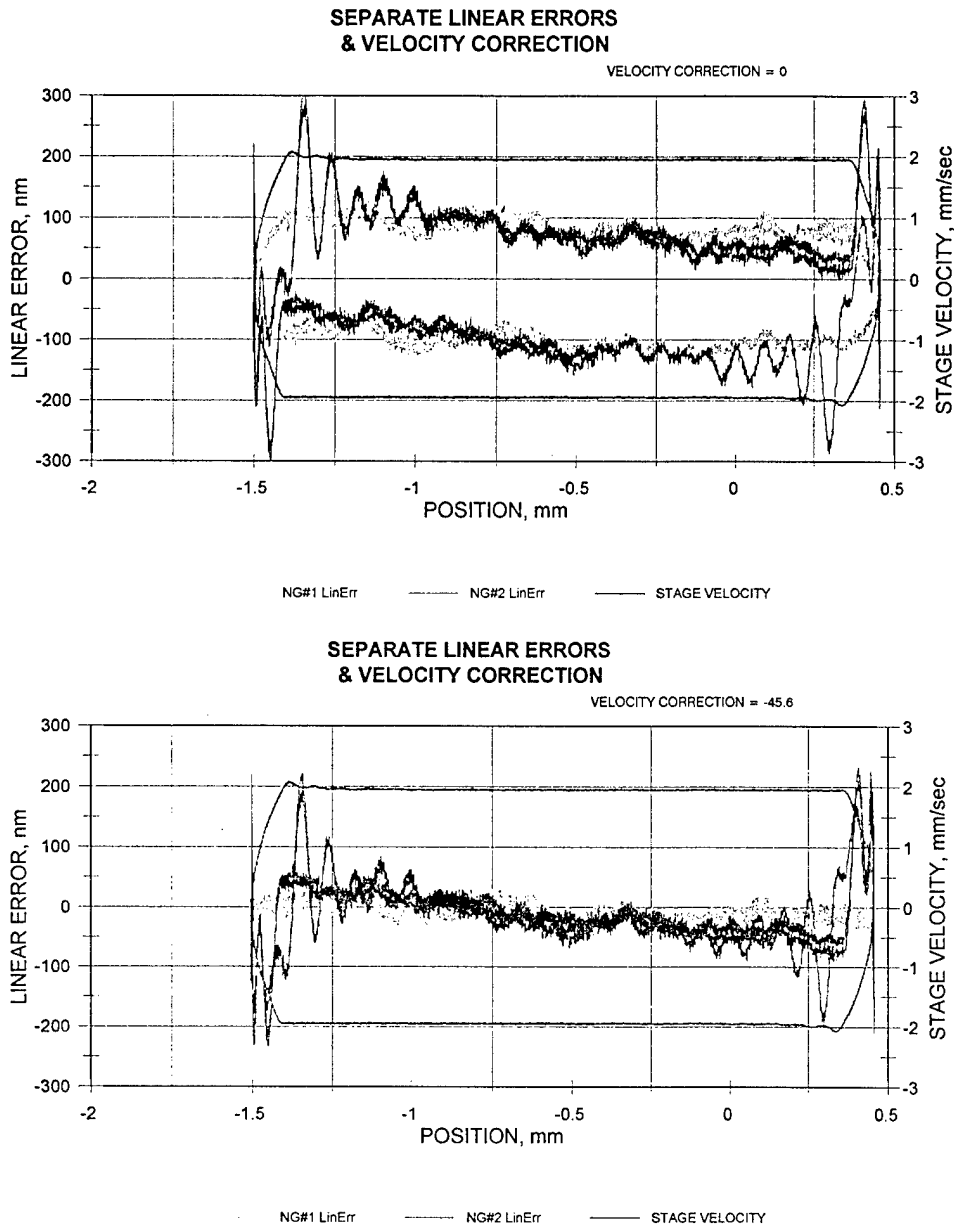


Figure 18 Linear Error Measurements & Latency Error

Because of the finite time between the time at which the stage arrives at any given position, the the time at which that position is reported (data latency), measurements must be corrected for the stage velocity. In the upper set of plots there has been now velocity correction, and as a result there is a marked hysteresis in the measurements. In the lower plot, a velocity correction of $-45.6 \text{ nm}/(\text{mm}/\text{sec})$ has been applied, which just about eliminates this hysteresis.

2.3 Hand-off Performance

Significant technical effort was directed toward the development of hand-off technology. One of the contract goals was to provide accurate coverage over an area of 300mm x 300mm. Our early metrology work led us to the conclusion that the most accurate NanoGrid performance is achieved when using a grid manufactured on an ETEC ALTA machine. Unfortunately, these machines are limited to producing grids at 175mm x 175mm. We sought to demonstrate the ability to extend the range of coverage by handing off a single grid to multiple sensor heads.

To understand the issues in a handoff design, one must understand how the Hi Res Processor measures the position of a grid relative to the sensor.

For each axis, the sensor outputs three analog signals (R, S, T) of the form:

$$R = A + B \cos \phi$$

$$S = A + B \cos(\phi + 120)$$

$$T = A + B \cos(\phi - 120)$$

The phase (ϕ) changes 360 degrees for a change in grid position of 5 micrometers. In the phase processor, these signals are amplified and converted to two signals (X, Y)

$$X = C \cos \phi$$

$$Y = C \sin \phi$$

These signals are digitized and applied to a coordinate transformation I.C. which outputs a 14 bit word proportional to ϕ .

The 14 bit word is applied to an unwrapping circuit which detects the transition between 0 and 360 degrees. The unwrapping circuit controls an up/ down counter that generates an 18 bit word. The counter output is combined with the phase word to form a 32 bit word proportional to the grid position relative to the sensor. Timing and control circuits synchronized to a 15 MHz clock control the digital circuits. The phase words are updated at a 375 KHz rate. An output clock signal (YCLK) is provided to indicate when the output position words are updated.

Handoff Issues

The issues addressed by the preliminary design are:

1. At what point in the processor circuit is the handoff done.
2. How to determine when to handoff.
3. How to account for the discontinuity in the phase word caused by the handoff.
4. Errors caused by the handoff.

Handoff Circuit

In the preliminary design, the handoff is done at the input to the processor. This requires only one set of processor circuits per axis.

Handoff Control

The preliminary design will use the position word to determine which sensor will be used to measure the grid position. This requires the processor counter to be initialized at a known grid position.

Handoff Algorithm

The handoff algorithm compensates for the discontinuity between sensors by adding a constant to the phase word before it is applied to the unwrapping circuit. This constant is adjusted whenever handoff occurs.

The algorithm stores 4 consecutive phase words, 3 from the old sensor, one from the new sensor. If we label these samples as A, B, C, D, the new constant K is calculated as follows:

$$K = K + (B - A) + (C - D).$$

The first term accounts for the velocity of the grid when handoff occurs. The second term accounts for the offset between sensors. Table 1 is a typical calculation.

Table 1

SAMPLE	PHASE	K	OUTPUT
A	15	8	23
B	21	8	29
C	27	8	35
D	48	-7	41

$$K = 8 + (21 - 15) + (27 - 48) = -7$$

Implementation

The algorithm, for a single axis, will be tested using an existing Hi Res Processor pcb. The handoff algorithm has been implemented in schematic form that will be programmed into a Xilinx Field Programmable Gate Array (FPGA). The FPGA is mounted on a small pcb that mates to a 40 pin connector on the Hi Res Processor. (This pcb was designed for another project). The Hi Res Processor also has an analog multiplexer that will be used for the handoff. The FPGA also uses the 15 MHZ clock and YCLK to generate the timing signals that control the latches that store the phase words.

Performance Predictions

The errors associated with the handoff circuit are caused by the sensor signals; interpolation errors, and Noise Equivalent Displacement (NEDX) errors associated with the detector shot noise and laser relative intensity noise. Another possible error source is acceleration induced errors.

Interpolation errors are specified as less than 10 nanometers p-p. NEDX errors are about 2 nanometers p-p. Acceleration induced errors are negligible.

Performance

The circuit was built and tested with simulated signals. Data were logged with a Hewlett-Packard Logic analyzer. No discontinuities or accumulated errors were noted after dozens of repeated hand-offs. The circuit was never tested in a live system.

2.4 Interpolation Error at High Speed

When this program began, we were aware that our sensor head interpolation error may increase with velocity. This was due to unequal bandwidth roll-off of the R, S and T amplifiers. During this program we re-designed the pre-amplifier assembly to allow precise matching of the R, S and T bandwidths. The result is a very flat response to velocity out to 400mm/sec. We were unable to test beyond this due to limits in our data

acquisition. To date we have tested eight sensors with the new configuration and all have behaved as expected. Test data is summarized in Figure 2.5.1 and Table 2.5.1.

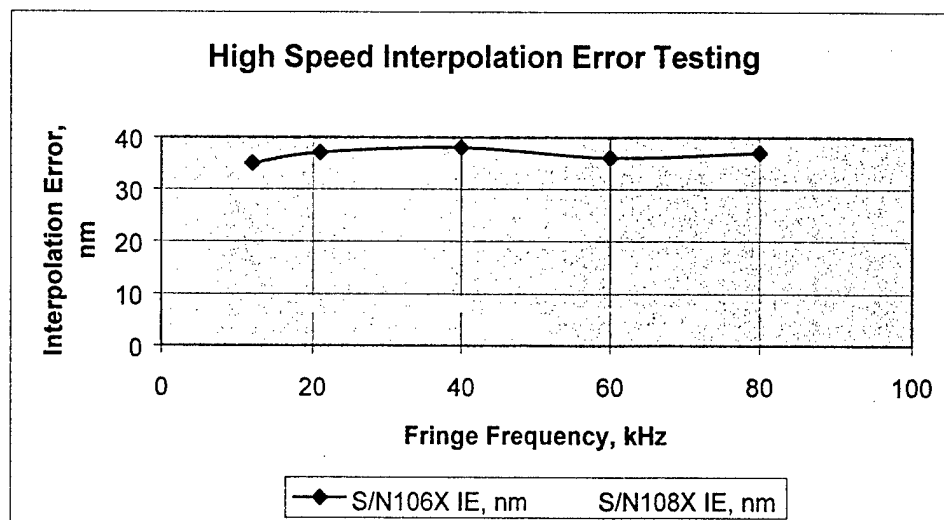


Table 2 High Speed Interpolation Error Testing.

S/N	75mm/sec (nm)	225mm/sec (nm)	DELTA (nm)
107Y	9	21	12
107X	7	9	2
117Y	14	24	10
117X	21	22.5	1.5
119Y	20	26.5	6.5
119X	10.4	10.8	0.4
120Y	10.1	17	6.9
120X	7.3	9.3	2
112Y	28	29.3	1.3
112X	25	22.7	-2.3
114Y	22.4	29.2	6.8
114X	9.9	10.2	0.3
Delta Avg			4.0nm
X-axis Delta Avg			0.7nm
Y-Axis Delta Avg			7.3

2.3 Integration with a Motion Controller

Obviously, a fundamental requirement for any metrology system is its ability to be integrated with a motion control system. For this program, we chose to integrate the sensor with a Delta-Tau motion controller, using the parallel word input option to take best advantage of our processor speed. We were able to close the loop at 40HZ before the program was ended. Final design considerations are presented below.

DIR Document DIR-136 Phase II Motion Requirements provides the requirements for the motion control design. The table below shows a summary of the upper and lower stage move parameters that support these requirements.

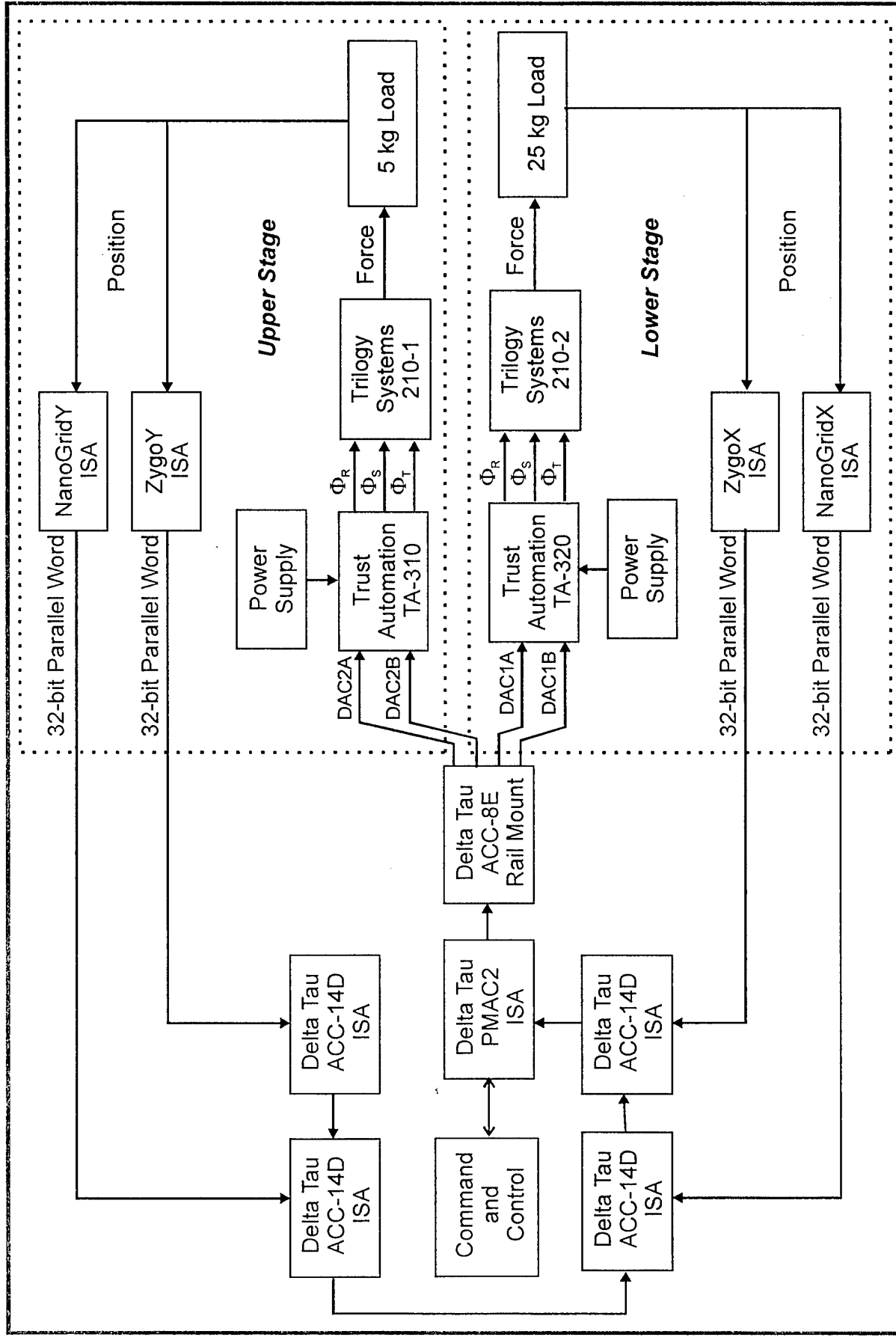
Table 1 – Stage Move Parameters		
Parameter	Lower Stage	Upper Stage
Axis	X	Y
Load	25 kg (55 lbs)	5 kg (11 lbs)
Travel distance	300 mm	300 mm
Move time	0.55 s	0.40 s
Rest time	3.0 s	3.0 s
Duty cycle	~15 %	~12 %
Maximum acceleration	4 m/s ²	10 m/s ²
Maximum velocity	1 m/s	1 m/s
Maximum velocity time	50 ms	200 ms
Maximum velocity travel	50 mm	200 mm
Motor	Trilogy Systems 210-2*	Trilogy System 210-1*
- Peak force	101.8 N	51.6 N
- Continuous force	37.9 N	12.5 N
Amplifier	Trust Automation TA320*	Trust Automation TA310*
- Peak current	3.8 A	3.9 A
- Continuous current	1.4 A	0.94 A
- Peak voltage	71.4 V	36.2 V

*Additional hardware requirements

1. Lower stage motor requires additional 5.4 inches magnetic track
2. Upper stage motor requires additional 5.4 inches magnetic track
3. Lower stage requires Trust Automation TA320 linear amplifier and a power supply
4. Upper stage requires Trust Automation TA310 linear amplifier and a power supply
 - Note that the Western Servo BLH-S1-4/8 linear amplifier would also support this move

Detailed Design

The figure below shows a block diagram of the hardware for each axis of the stage. The sections following the diagram summarize the detailed design for each piece of hardware.



Delta Tau PMAC2 Motion Controller

The Delta Tau PMAC2 motion controller will be used to provide motion control in both axes of the stage system shown in the figure. Note the system will be designed to accommodate encoder feedback from either one of two different sources for each axis. Note also that the terms *PMAC*, *PMAC2*, and *Turbo PMAC2* are used interchangeably in this document.

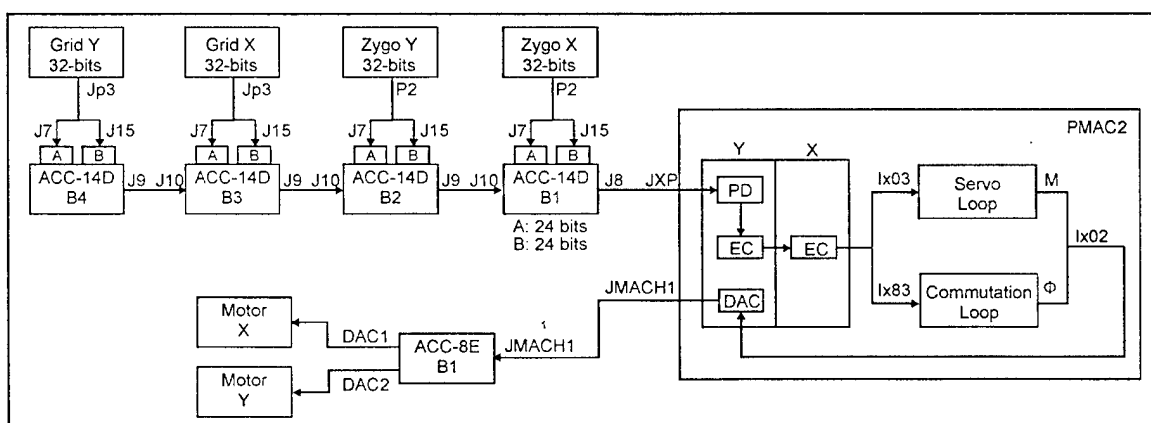
Note on Hardware & Software Setup of the PMAC2

The PMAC2 is shipped from the factory with hardware and software defaults set up for the most common application types. Thus, the PMAC2 will have to be configured for this specific application. For the hardware, the PMAC2 (and other accompanying accessory boards) contain many jumpers, called E-points, which are used to customize the hardware setup for an application. Similarly, for the software, PMAC2 has a large set of initialization parameters (I-variables) that set up the functionality of a card for a particular application. Once set up, these variables are stored in non-volatile EAROM memory (using a **SAVE** command) so that card remains properly configured; PMAC2 loads the EAROM I-variables into RAM upon power-up.

Setup, tuning, and programming of the PMAC2 is done through the PMAC executive program PEWIN32 in conjunction with the P2 Turbo Setup program. PEWIN32 is a windows based program that provides a number of features, of which some are:

- A terminal window for communication between the user and PMAC2
- Interface for setup and handling PMAC2's I, P, Q, and M-variables
- Position, velocity, and following error display for all motors
- Several methods for tuning PMAC2
- Data gathering and plotting capabilities
- Real-time status feedback
- And many more...

The following figure provides a detailed diagram of the PMAC2 setup for incorporating parallel feedback from up to four encoders and then generating position and commutation closed-loop commands.



PMAC2 Memory Locations and Mapping for Parallel Feedback

The following tables detail the memory locations for the parallel feedback data, the memory mapping for reading the encoder data into the encoder conversion table, the encoder conversion table output memory locations used for servo position update, and the memory locations used for commutation. As shown in the previous figure, parallel data is brought into PMAC2 via the ACC-14D board.

ACC-14D Source Register Addresses		
Encoder	ACC-14D Input	Y-Address
Zygo X	Board 1 Port A (J7)	\$78A00
Zygo Y	Board 2 Port A (J7)	\$78B00
NanoGrid X	Board 3 Port A (J7)	\$78C00
NanoGrid Y	Board 4 Port A (J7)	\$78D00

Encoder Conversion Table INPUT Source Register Values			
I-Variable	Y-Address	Value	Comment
I8000	\$003501	\$278A00	Zygo X encoder data location; normal (shifted processing)
I8001	\$003502	\$018000	Use 24-bits; start bit 0
I8002	\$003503	\$278B00	Zygo Y encoder data location; normal (shifted processing)
I8003	\$003504	\$018000	Use 24-bits; start bit 0
I8004	\$003505	\$278C00	NanoGrid X encoder data location; normal (shifted proc)
I8005	\$003506	\$018000	Use 24-bits; start bit 0
I8006	\$003507	\$278D00	NanoGrid Y encoder data location; normal (shifted proc)
I8007	\$003508	\$018000	Use 24-bits; start bit 0

Encoder Conversion Table OUTPUT Source Register Values	
X-Address	Comment
\$003502	Zygo X encoder converted data
\$003504	Zygo Y encoder converted data
\$003506	NanoGrid X encoder converted data
\$003508	NanoGrid Y encoder converted data

The following tables provide the I-variable setup information for each of the encoder types.

I-Variable Position & Commutation Setup for Zygo Feedback		
I-Variable	Value	Comment
I103	\$003502	Position feedback information for lower stage (X-axis)
I183	\$78A00	Commutation unfiltered position information for lower stage motor
I203	\$003504	Position feedback information for upper stage (Y-axis)
I283	\$78B00	Commutation unfiltered position information for upper stage motor

I-Variable Position & Commutation Setup for NanoGrid Feedback		
I-Variable	Value	Comment
I103	\$003506	Position feedback information for lower stage (X-axis)
I183	\$78C00	Commutation unfiltered position information for lower stage motor
I203	\$003508	Position feedback information for upper stage (Y-axis)
I283	\$78D00	Commutation unfiltered position information for upper stage motor

The following table provides the jumper setup and address locations for the four ACC-14D boards.

ACC-14 D		Jumpers						PMAC2 Address		
Input	E12	E13	E14	E15	E16	E17	E20	J7 (Port A)	J15 (Port B)	Control Word
B1 - Zygo X	On	Off	Off	Off	Off	Off	Off	\$78A00	\$78A01	\$78A03
B2 - Zygo Y	Off	On	Off	Off	Off	Off	On	\$78B00	\$78B02	\$78B03
B3 - NG X	Off	Off	On	Off	Off	Off	On	\$78C00	\$78C03	\$78C03
B4 - NG Y	Off	Off	Off	On	Off	Off	On	\$78D00	\$78D04	\$78D03

The ACC-14 D must be software configured each time it is powered up to set which bytes are inputs and which are outputs. This is done by writing to the control word (base address + 3 as shown in the above table) the value \$93929B, which configures all 6 bytes for input. This will be done with a PLC program that automatically writes the control words at power-up and then disables itself.

PMAC2 Motor Setup

The PMAC2 will be setup for controlling the motors in analog sine-wave command mode, i.e. sinusoidal commutation. This will provide smoother operation over the currently used torque-mode commutation with hall feedback as well as totally eliminating any cogging at low speeds. The machine interface ports on the PMAC2 are connected to the motor via interface board ACC-8E (100-pin header connector JMACHn on PMAC2 to the mating JMACH connector on ACC-8E), which provides dual analog outputs per axis for velocity, torque, or sine-wave commands. The ACC-8E board also provides encoder (quadrature), hall, and optically isolated flag feedback (e.g. homing, +/- limits, etc.). Note that this board will not be used for encoder feedback since the encoders provide a parallel output, which will be read in using a different board (ACC-14D).

ACC-8E Specific Information

The majority of the ACC-8E information will be outlined in the following sections. This section provides some generic setup information.

- ACC-8E requires a separate +/- 12V to +/- 15V supply via terminal block TB1, which should be isolated from the PMAC2's 5V supply (which is used to power the digital circuitry on ACC-8E).
- The 5V digital reference is brought over from the PMAC2.
- The main input flags, e.g. Home, Positive Limit, Negative limit, General User, for each axis are brought in on terminal blocks TB4 and TB6. The flags are provided with optical isolation and sinking or sourcing capability to 24V. The default for the limit switch is true low for normal operation (a high indicates an overtravel limit). If no limits are used, either hardwire directly to ground or set the limit switch to true high in software (a low indicates a fault).

PMAC2 Sine-Wave Output Control

The PMAC2 will be setup for sinusoidal commutation. In this mode, the PMAC2 provides two phase current commands to the Trust Automation amplifiers via DAC's on the ACC-8E interface board. The third motor command is then derived by the amplifier from the PMAC2 signals. The PMAC2 also provides amplifier enable outputs and amplifier fault detection.

- Analog outputs

The analog outputs are differential and have a range of +/- 10V. The analog outputs can be obtained either via the 16-pin header connector J4 (both motor outputs) or 12-pin terminal block connectors TB5 (DAC1) and TB7 (DAC2). The analog outputs are clamped at 0V in the

event of loss of on-board power, the amplifier enable signal goes false, or by tripping the PMAC2's watchdog timer.

- **Amplifier enable outputs**

The PMAC2 provides amplifier enable outputs via either connector J4 or connectors TB5 and TB7. The Trust Automation amplifiers require TTL 0 enable signals, which implies jumper E2 (E5) is open and E4 (E7) is closed for amplifier 1 (2). Also, since the amplifier enable input is single ended we will tie the negative side of the amplifier enable output to common by closing jumper E3 (E6) for amplifier 1 (2). The table below shows the ACC-8E jumper settings.

ACC-8E Jumper Settings		
Jumper	Description	Setting
E1	Provides digital 5V reference via JMACH from PMAC2 when closed	Closed
E2	Provides for normally closed (open when enabled) for amplifier 1 when closed	Open
E3	Ties amplifier enable minus to analog common for amplifier 1 when closed	Closed
E4	Provides for normally open (closed when enabled) for amplifier 1 when closed	Closed
E5	Provides for normally closed (open when enabled) for amplifier 2 when closed	Open
E6	Ties amplifier enable minus to analog common for amplifier 2 when closed	Closed
E7	Provides for normally open (closed when enabled) for amplifier 2 when closed	Closed

- **Amplifier fault inputs**

The PMAC2 provides amplifier fault inputs to ACC-8E via connectors TB5 and TB7. The PMAC2 default for the fault I/O is enabled (on) with a low-true input indicating a fault. The input is pulled high internally so if no fault line from the amplifier is connected to the input the motor will still operate.

Analog sine-wave commutation global variable setup

I-variables I7000 – I7005 set up PMAC2 global variables for proper operation in sine-wave commutation mode. Every phase clock cycle, PMAC2 performs the commutation for each motor, at the rate set by variables I7m00 (MaxPhaseFreq) and I7m01 (PhaseFreq). For sine wave mode, I7m00 = I7m01. Furthermore, for parallel data, Delta Tau recommends setting the servo frequency (I7m02 – ServoFreq) equal to the PhaseFreq. Special consideration needs to be considered when setting the servo frequency.

In general, there are three considerations that enter into properly setting PMAC up to perform sinusoidal commutation with parallel feedback: 1) required servo rate (frequency), 2) maximum allowable counts per commutation cycle, and 3) saturation of the velocity filter buffer. Basically, as will be shown below, the best possible solution will be to provide a ~10 nm LSB from the encoders to the PMAC ACC-14 board.

1. For parallel data, the sample rate (ServoFreq) must be sufficient to read the input word fast enough that less than half of the range of the input word has been covered since the last sample (effectively the Nyquist limit). Parallel data is brought into PMAC2 via the ACC-14 I/O board, which latches its inputs every servo cycle. The number of bits of the parallel word PMAC processes is 24. Two modes of operation exist, normal (shifted) and unshifted. The Darpa system will be setup for normal shifted mode. In this mode, the maximum number of

bits the PMAC2 can use is 19, which yields a maximum velocity of $2^{18} - 1$, or 262,143 counts per servo cycle. Thus, for a 1 m/s velocity, the servo rate needs to be at least:

$$\text{ServoRate} \leq \frac{2^{18} - 1}{10^9} \times \text{LSB}_{nm}$$

The table below shows the minimum required servo rate to support the encoder (either the Zygo or NanoGrid) with a resolution of 9.6 nm.

LSB (nm)	Servo Rate (ms)	ServoFreq (Hz)
9.6 nm	2.52	~400

The default rate for the PMAC2 is on the order of 9 kHz, which will be sufficient for this mode of operation. Note that PMAC "looks" for a servo count in the LSB+5 bit of the 24-bit word out of the encoder conversion table, which means the LSB of the feedback device is 1/32 of a PMAC "count". Finally, note also that in this mode 2's-complement data will not be an issue.

2. In order to perform sinusoidal commutation, PMAC needs to know the number of counts per commutation cycle, which is obtained by dividing I-variable Ix71 by Ix70. For linear motors, Ix70 defaults to 1, and the counts/cycle value is contained in Ix71, which has a maximum value of 8,388,607. The Trilogy motor has a 2.4 in commutation cycle, which results in

$$\text{Ix71} = \frac{2.4 \text{ in}}{\text{comm cycle}} \times \frac{1 \text{ count}}{9.6 \text{ nm}} \times \frac{1 \text{ nm}}{1 \times 10^{-7}} \times \frac{2.54 \text{ cm}}{\text{inch}} = 6,350,000 \frac{\text{counts}}{\text{comm cycle}},$$

which is less than maximum allowable value.

3. Finally, PMAC's velocity filter will saturate when the velocity (in counts/sec) multiplied by Ix08 exceeds 768,000,000. For this mode of operation (PMAC count = 9.6 nm and 1 m/s maximum velocity),

$$\text{Ix08} \leq \frac{768,000,000}{\frac{1 \text{ count}}{9.6 \text{ nm}} \times \frac{10^9 \text{ nm}}{\text{s}}} = 7.37.$$

The table below provides a summary of the global variable setup for the PMAC2.

Parameter	Value	Comment
I7000	6527	Default – MaxPhaseFreq = 9.0346 kHz
I7001	0	Default – PhaseFreq = MaxPhaseFreq = 9.0346 kHz
I7002	0	ServoFreq = PhaseFreq = 9.0346 kHz
I10	928497	Amount of time between servo interrupts – function of I7000, I7001, I7002

Analog sine-wave commutation per-channel variable setup

First, we need to define the mapping of channels. Our Turbo PMAC2 card provides for up to 4 machine interface channels, which are implemented using the ACC-8E h/w option. Each ACC-8E can configure two channels. We are using one ACC-8E board since we only need to control two axes. Thus, we will define channel 1 as the machine interface for the lower stage (x) and channel 2 as the machine interface for the upper stage (y). Furthermore, we will define the lower stage as motor 1 and the upper stage as motor 2, resulting in the situation where the channel number is equivalent to the motor number (PMAC convention uses the 10's digit to identify channel number in the channel I-variables and the 100's digit to identify motor number in the motor I-variables).

Parameter	Value	Comment
I7016	1	Sets channel 1 outputs A & B for DAC mode
I7026	1	Sets channel 2 outputs A & B for DAC mode

Analog sine-wave commutation motor variable setup

Several I-variables need to be set up for each motor x to enable and configure the sine-wave output. The following table provides the required I-variable settings for the ARPA stage system.

I-Variable	Value	Comment
I100	1	Sets motor 1 active
I200	1	Sets motor 2 active
I101	1	PMAC commutates motor 1
I201	1	PMAC commutates motor 2
I108	1	Position scale factor for motor 1; see previous section
I208	1	Position scale factor for motor 2; see previous section
I109	1	Velocity loop scale factor for motor 1; same value as Ix08
I209	1	Velocity loop scale factor for motor 2; same value as Ix08
I170	1	Linear motor has one commutation cycle for each motor travel cycle; Motor 1
I270	1	Same as above; Motor 2
I171	6350000	Number of counts per commutation cycle (see previous section)
I271	6350000	Same as above; Motor 2*
I172	683 or 1365	Phase commutation angle; either +/- 120 – set exp (wiring dep); Motor 1
I272	683 or 1385	Same as above; Motor 2
I173	1000-4000	Phasing search output; Trilogy value; Motor 1
I273	1000-4000	Same as above; Motor 2
I174	5	Phase search command time in servo cycles; Trilogy value; Motor 1
I274	5	Same as above; Motor 2
I180	1	Perform two-guess phasing method at power-up/reset; Motor 1
I280	1	Same as above; Motor 2

PMAC2 Integration Tests

In general, basic operation of the commutation outputs, servo drive, motor, and feedback will be established by using output offset variable Ix29 and Ix79 to force current directly into the particular phases and drive the motor like a stepper motor. This test will verify:

- That output voltages are obtained from the ACC-8E board

- That currents flow in the phase of the motor
- That currents cause the motor to lock into position
- Proper polarity of the Ix72 commutation phase angle

The details of this test will be outlined in a test procedure.

Absolute Encoder Latching for Parallel Inputs

As shown previously, parallel data is brought into PMAC2 from the encoder outputs via ACC-14D. It is extremely important to properly latch the encoder data to prevent PMAC2 from reading the encoder data during an encoder transition. The ACC-14D allows for 5 different latching schemes; our system will use Method 3, which requires a self-latching encoder. In this method the encoder outputs a signal indicating it is latched, which is brought into the ACC-14D via the ICLK inputs (connector J7 and/or J15). The ACC-14D will then only strobe (latch) the encoder inputs on the falling edge of the ICLK only when the servo clock is low, preventing misreads. The table below shows the ACC-14D jumper settings for this configuration.

Required Signal	E5/E6	E8/E9	E21/E22
High ICLK means latched	On	Off	Don't care

PMAC2 Detailed Design Open Issues

1. PMAC2 interface communication with data acquisition software
2. Setup PMAC2 to trigger data acquisition
3. Motor home sensors and homing routine
4. Axis overtravel limit information and PMAC2 setup
5. Calibration trajectories and other test trajectories

TA320 Servo Amplifier

The lower axis of the stage will use the Trust Automation TA320 servo amplifier to drive the LM210-1 Trilogy Motor. The table below shows the amplifier performance for the required design.

Parameter	Requirement	Amplifier Capability
Peak Current	3.8 A	12 A
Continuous Current	1.42 A	6 A
Peak Voltage	71.4 V	120 V

The amplifier also has a number of switches. The table below shows the settings for switch S2 to support this application, along with definitions.

TA320 Switch S2 settings		
Switch #	Setting	Comment
1	Down	Aux Gnd tied to GND
2	Down	Amplifier supplies 5V
3	Down	Torque (current) mode
4	Up	Differential command input
5	Up	T-phase internally derived from input R & S phases
6	Up	Amplifier fault generates a low input to PMAC
7	Down	Should not matter with differential input on P5
8	Down	Should not matter with differential input on P5
9	Down	Should not matter with differential input on P5
10	Up	Sinusoidal commutation

The current limit is set via switch S1 and will be set to 6 A as required by the requirement, which is position 2.

TA310 Servo Amplifier

The upper axis of the stage will use the Trust Automation TA320 servo amplifier to drive the LM210-1 Trilogy Motor. The table below shows the amplifier performance for the required design.

Parameter	Requirement	Amplifier Capability
Peak Current	3.9 A	8 A
Continuous Current	0.9 A	4 A
Peak Voltage	36.2 V	48 V

The amplifier also has a number of switches. The table below shows the settings for switch S1 to support this application, along with definitions.

TA310 Switch S1 settings		
Switch #	Setting	Comment
1	Down	Aux Gnd tied to GND
2	Down	Amplifier supplies 5V
3	Up	Amplifier fault generates a low input to PMAC
4	Down	Current Mode
5	Down	Along with switch 6 sets current limit at 6 A
6	Up	Along with switch 5 sets current limit at 6 A
7	Up	Sinusoidal commutation
8	Up	120 Hall commutation (Should not matter for sinusoidal commutation)

LM210-2 Linear Motors

The lower axis of the stage will use the LM210-2 Trilogy Motor. The table below shows the LM210-2 motor performance for the required design.

Parameter	Requirement	Maximum Capability
Peak Motor Force	101.8 N	133.4 N
Continuous Force	37.9 N	62.3 N
Temperature Rise	29.2 C	100 C

Wiring for sinusoidal commutation is very simple; all that needs to be done is to hook-up the three respective leads to the servo amplifier in any order. The phase angle direction will be set up in PMAC software via I172.

LM210-1 Linear Motors

The upper axis of the stage will use the LM210-1 Trilogy Motor. The table below shows the LM210-1 motor performance for the required design.

Parameter	Requirement	Maximum Capability
Peak Motor Force	51.6 N	66.7 N
Continuous Force	12.5 N	28.9 N
Temperature Rise	29.2 C	100 C

Wiring for sinusoidal commutation is very simple; all that needs to be done is to hook-up the three respective leads to the servo amplifier in any order. The phase angle direction will be set up in PMAC software via I272.

Zygo Encoder

A Zygo ZMI 2000 measurement system will provide position information for the lower stage (x-axis). The Zygo outputs a 36-bit parallel word with 0.31 nm resolution; a range of travel over ± 10 m; and a maximum velocity of 2.1 m/s. The PMAC will use the lower 24-bits of this word to update the position loop. The fact that it is 2's-complement is no issue to the PMAC processing of the data. The position information will be provided to PMAC via the Zygo board 32-bit P2 bus, along with signal indicating when it is latched (High when latched).

NanoGrid Encoder

A NanoGrid two-axis encoder will provide position information for both the lower and upper stage axes. The NanoGrid outputs a 32-bit parallel position word with a 0.305 nm resolution; a range of travel up to 1.3 m; and a maximum velocity of up to 1.4 m/s. The NanoGrid will be setup in a similar fashion as the Zygo encoder. Again, the 2's-complement output will not be an issue for PMAC and the lower 24-bits will be used along with the latching signal (High when latched) to update the position loop.

Section 3. Applying Planar Encoder Technology to Next Generation Lithography

The initial intent of this program was for OPTRA to exploit the advantages of the NanoGrid metrology and develop a stage for use in next generation lithography machines. Extensive discussions with both stage manufacturers (Anorad, Dover) and potential customers (ASML, Nikon, Canon, SVGL, UltraTech) led us to the conclusion that the major players in the lithography field all make their own stages and will continue to do so. Furthermore, customers who do buy stage plus metrology systems generally require a custom system specifically designed for their application. Given this, we felt it best to use this program to develop a versatile stage metrology system, independent of a specific stage configuration. This thought process has been documented more fully in the monthly reports filed to date.

The metrology requirements for a NGL system are derived from the stage requirements, which are in turn derived from the International Technology Roadmap for Semiconductors. The parameters that particularly drive the stage-metrology requirements are wafer format, throughput and critical dimension node, which translate into specifications for range of travel, speed and accuracy respectively. The following table lists the metrology requirements for a NGL system and the demonstrated NanoGrid capability.

Table 3. NGL Stage Metrology Requirements and NanoGrid Capability

Parameter	Specification	NanoGrid Capability
Range of Travel	>300mm × 300mm	Unlimited with multiple sensor hand-off
Speed	>1 m/sec	1.5 m/sec
Resolution	<1nm	1.0 nm RMS
Repeatability	±2.5nm	±2.5nm
Accuracy (uncorrected)	±100nm	±75nm

A casual assessment of Table 1 leads one to believe that the NanoGrid can in fact satisfy the metrology requirements for NGL. Indeed, we at OPTRA firmly held this conviction at the inception of this program, especially when one added the additional benefits of reduced turbulence errors, reduced Abbe errors and lower cost. As the program progressed, however, we learned that there were challenges associated with using grid technology that greatly diminished its appeal to the lithography community.

One development, in particular, was the emergence of SCALPEL, DUV and EUV technologies as the lead technologies for NGL. Once implemented, all of these technologies will operate in a vacuum, thus negating the need for reduced sensitivity to turbulence. During our visit to the DARPA Advanced Lithography Program Review in April 2000, it was suggested by several people that stage metrology was a non-issue. The laser interferometers worked fine – offering 0.6nm resolution at speeds of 2meters/sec – and there were bigger issues to be addressed. SSG considered using the NanoGrid in their SCALPEL stage development, but felt it was too risky to change from the established laser interferometer metrology.

Finally, there is the issue of range of travel. As we have mentioned earlier in this report, we were unable to develop a technique for the fabrication of 300+mm grids of lithography-grade accuracy. Even our most accurate grids contain high spatial

frequency errors that make them unfit for use in lithography applications. We have demonstrated the feasibility of multiple sensor hand-off which would theoretically infinitely extend the range of travel – just keep adding more sensor heads. This is an expensive approach, however, that lessens that Abbe error advantage of a single sensor. Given all this, we reluctantly recognize that the NanoGrid is not a good fit for NGL stage metrology.

The news is not all bleak, however. As of this writing, we have successfully fielded NanoGrid sensors in the GSI M430 memory repair machine. This machine promises to dramatically increase the speed at which DRAM chips can be fabricated. Our success in this area is due largely to insights about grid metrology that were developed under this contract. We continue to look for other niche applications within the lithography marketplace that may more easily accept our encoder system. In particular, we are looking for application areas that are suited for limited stroke planar stages. Memory repair is one area we have already identified, we continue to explore other chip-level applications such as wire-bonding and optical inspection.

Section 4. The NanoGage

We have developed a compact optical sensor to measure the longitudinal position of a specularly or diffusely reflective surface. The measurement is based on triangulation, but uses a bi-cell detector instead of the more usual position-sensing detector. A beam from a laser diode is focused to a spot with a diameter of approximately 100 microns on the target surface at a nominal distance of 3 mm from the sensor. Longitudinal displacement of the target surface is sensed with a resolution on the order of 20 nm, with a nearly linear response over a range of about ± 50 microns, and a capture range (i.e. the range over which a nulling servo can be operated) of about ± 400 microns.

Measurement principle

The NanoGage is a folded optical triangulation sensor based on a bi-cell detector and precisely controlled focal spot size. The optical layout is illustrated in figure 1 below. High resolution results from the use of a small focal spot size in conjunction with an AGC-controlled laser diode source and a bi-cell detector with a 10 micron gap. Individual sensors can be corrected to achieve 0.1% or better linearity over their specified measurement range. Automatic Gain Control (AGC) provides uniform response for a wide range of surface reflectivities. Unlike triangulation sensors based on position-sensing detectors, in which the linear range is set by the detector size, the NanoGage sensor has a linear range that is defined by its focal spot size. This allows for very high resolution in a compact and robust package.

Good linearity is achieved by tailoring of the lens aberrations to give a uniform circular focal spot; final linearity is accomplished based on individual sensor calibration and correction.

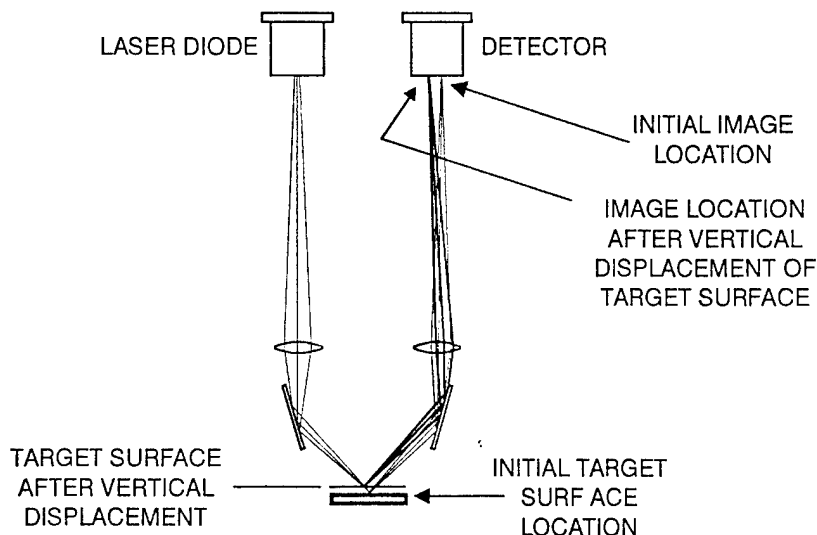


Fig.1 – NanoGage optical layout

Because of the oblique illuminating and viewing angles, a vertical displacement of the target surface causes a lateral displacement of the laser image in the plane of the bi-cell detector. The signals from the bi-cell are processed to generate a voltage that is proportional to the vertical displacement. The fold mirrors allow the sensor to be configured as a compact cylinder.

Figure 2 shows how the (A-B) signal from the bi-cell detector is generated from the lateral displacement of the laser image on the detector, and figure 3 shows actual data from a NanoGage sensor calibration against a linear encoder with an accuracy of better than 20nm.

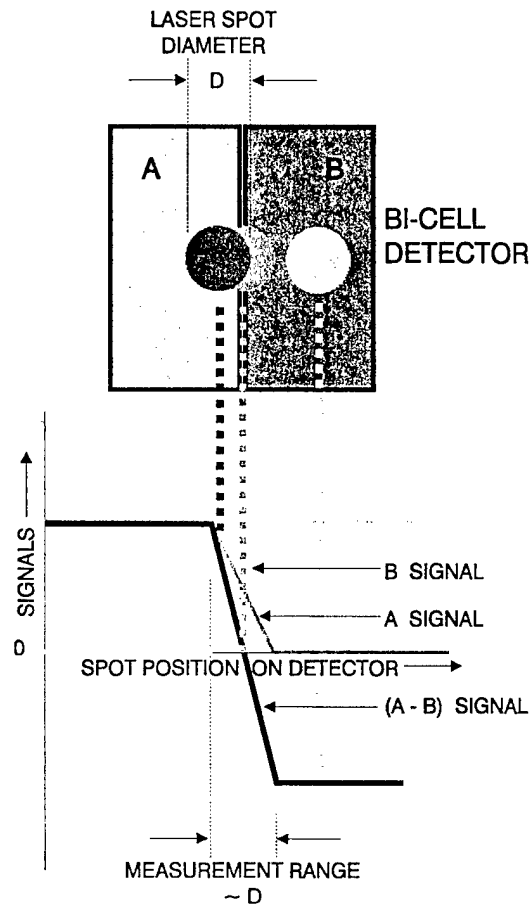


Fig - 2 Bi-cell detector signals

Each of the bi-cell elements generates a separate signal due to light from the laser spot image. As shown, the (A-B) plot is approximately linear with the spot location of the detector; the exact shape of the plot depends on the distribution of light within the laser spot. When the spot is on the dividing line between the two detector elements, the signals are balanced and $(A-B)=0$.

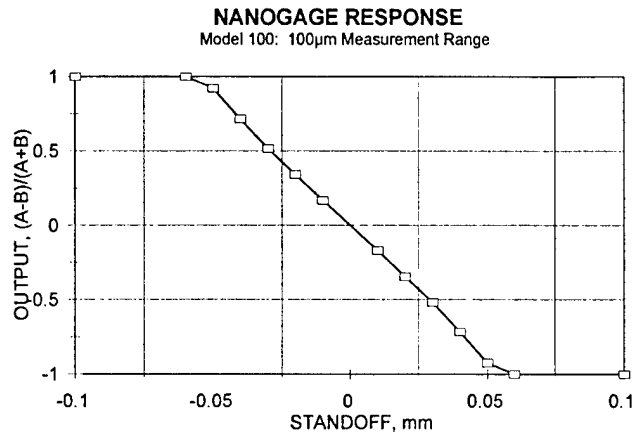


Fig – 3 NanoGage response curve

This is the uncorrected response curve for the NanoGage Model 100 sensor. The measurement range is 100 μ m and the resolution is 25nm. After automatic electronic correction, the linearity can be corrected to 0.1% over the full specified measurement range.

The NanoGage's high resolution is due to the large signals made possible by the use of a laser diode light source. The target surface may be a mirrored surface, an un-coated glass surface, or a finely ground or machined metal surface. Unlike a capacitance gage, the NanoGage does not require a conductive surface, and is largely immune to electrical or magnetic interference. The structure of the NanoGage is made of Invar to minimize the effects of temperature on the sensor performance.

Configuration of the NanoGage sensor

As shown in figure 4, the NanoGage sensor has a cylindrical form and is designed to be held in a 12.7mm diameter split-ring to provide a 3mm nominal standoff from the target surface. A detector preamp is contained in the sensor head and the remaining electronics are built into the rectangular connector at the other end of the cable. This connector has a built-in light to indicate the target position: the light is red when the target is out of range, yellow when it is within the capture range, and green when it is within a few microns of its null position (at the center of its range).

Within the small sensor head are a number of optical components, including a laser diode, bi-cell detector, GRIN lens, red filter, 3 lenses and two mirrors. The location of the GRIN lens can be adjusted to set the laser spot size on the target surface to any desired size; the size of this spot determines both the range of the measurement and its resolution.

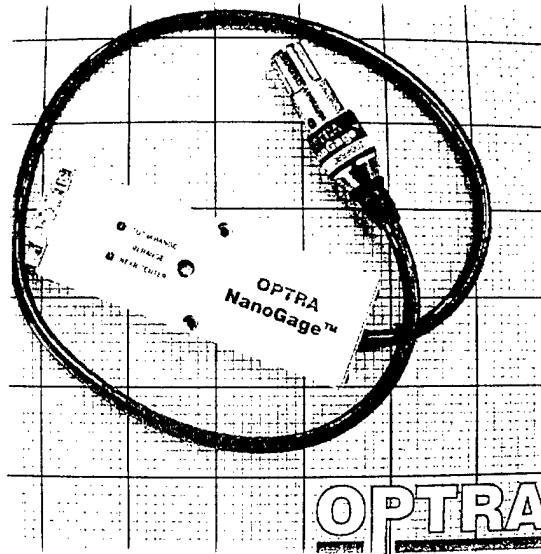


Fig – 4 NanoGage sensor & connector

The sensor is shown on 1" graph paper; the connector contains amplifiers and the laser diode drive circuit.

Sensor performance

The NanoGage offers accuracy and resolution comparable to a laser interferometer or high-end optical encoder, but in a much smaller package and at a far lower cost. The price to be paid for these economies is a limited measurement range (from 10 μ m to 500 μ m depending on the model).

In many applications, the NanoGage can replace a capacitance gage—offering following benefits:

- Low temperature sensitivity (Invar structure);
- High immunity to electric and/or magnetic fields;
- 3 mm standoff from the surface to be measured;
- Large capture range (range within which the sign of the signal tells in which direction the null position lies; typically ± 0.5 mm);
- High measurement bandwidth (1 MHz)—making the sensor well-suited for vibration analysis and high-speed run-out measurements; and
- Small measurement footprint (the measurement footprint is the area on the target surface that is involved in the measurement); the NanoGage measurement footprint ranges from 30 μ m diameter to 400 μ m diameter, depending on the choice of measurement range and resolution.

In addition to its measurement specifications, the NanoGage is compact and easy to mount for a wide variety of applications. The standard 12.7 mm diameter cylindrical surface also provides an easy way to adjust the initial standoff so that the desired measurement range is matched to the sensor's measurement range.

Applications

Servo position control

Since the NanoGage has an output signal that is proportional to the displacement of the target surface from its null position (where the [A-B] signal is zero), and a high bandwidth, it is well-suited as the sensor in a position servo-control system. For example, a linear stage may move from side to side as it traverses its length. This motion could readily be detected with a NanoGage, and servo controlled with the addition of a short-range displacement transducer (e.g. a piezoelectric transducer mounted on the stage slide. Such a servo control system could compensate for displacement amplitudes of $\pm 50\mu\text{m}$ with a resolution of 50nm. A similar arrangement could be employed to stabilize the height of a stage in order to maintain precise focus in a microlithography application.

Vibration measurement

Because of its linear response down to DC, and its high measurement bandwidth, the NanoGage is very well-suited for making longitudinal vibration measurements over a very wide range of frequencies. Note that the output of the NanoGage can either be directly fed to an electronic spectrum analyzer, or it can be digitized, saved as a text file, and later imported into a spreadsheet — where the Fourier analysis can be carried out, analyzed, and displayed.

Run-out measurements

Run-out measurements on rotating components allow the assessment of vibrations, eccentric mounting, and shaft wobble. These are critical issues in precision machinery, such as hard disc drives. The NanoGage can easily be used to make such measurements with very high precision, and for high rotational speeds. For example, at 9600rpm, the NanoGage's 100kHz frequency response allows an angular resolution of $\pm 0.5^\circ$ combined with a runout resolution of from 2.5nm to 125nm, depending on the model and its associated measurement range.

Surface profiling

Surface profiling is straightforward with a NanoGage and can be used with a wide variety of materials, ranging from glass and polished metals to plastics and paper.

Film thickness measurement

A pair of NanoGage sensors can do a very nice job of monitoring the thickness of materials either during fabrication, or as a quality control function. The sensors should be selected so that their measurement range corresponds to the full range of expected displacement between each sensor and the surface being measured. For transparent films, the two sensors should be slightly offset, to prevent light emitted by one sensor from being detected by the other sensor (a 2 or 3 mm offset is adequate).

Test Data

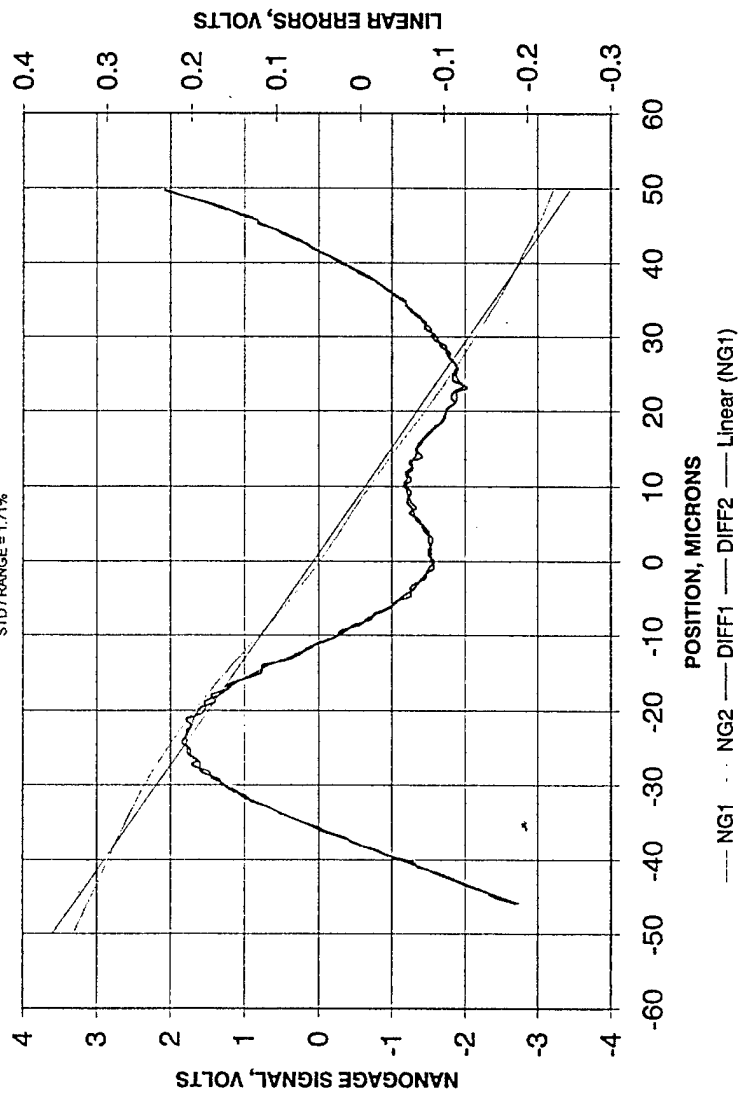
Ten prototype sensors were built and tested. the following charts summarize their performance.

Table 1: NanoGage Nominal Performance

Housing Number	Slope (mV/ μ m)	linear Range (μ m)	linearity ($\pm \mu$ m)	Capture Range ($\pm \mu$ m)	NE Δ X (nm)
P16	69.8	> 100	+1.5 / -2.6	+745 / -655	8.7
P17	53.1	> 100	+1.5 / -1.3	+720 / -700	12.6
P18	69.5	> 100	+3.4 / -3.6	+850 / -480	7.5
P19	72.2	> 100	+2.5 / -1.8	+460 / -860	9.7
P22	91.3	> 100	+5.3 / -2.9	+620 / -1390	9.5
P24	63.1	> 100	+2.2 / -1.7	+760 / -1080	11.25
P26	71.1	> 100	+2.5 / -1.7	+890 / -760	10.3
P27	68.5	> 100	+5.0 / -2.9	+780 / -890	10.9
P28	89.2	> 100	+4.7 / -3.5	+1230 / -870	8.4
P29	69.4	> 100	+1.6 / -1.6	+910 / -630	9.6

NanoGage # P18; Repeatability

PK-PK LINEAR ERROR / RANGE = 6.38% [4.47% OVER CENTRAL 80µm]
 STD / RANGE = 1.71%



APPENDIX A: Making an Error-Free Holographic LINEAR Encoder without Optics

Introduction

An error-free linear encoder is an encoder in which the pitch, or spacing between rulings, is constant from end to end. In the past, such encoders were either ruled in the same manner that diffraction gratings were ruled, or produced lithographically by reducing a large-scale drawing of the encoder and using a step-and-repeat process to make a long encoder. More recently, holography has offered an attractive alternative as a technique for making uniform encoders. In this technique, light from two coherent point sources is collimated and the resulting beams are made to overlap in the region where the holographic encoder is to be exposed (see figure 1). This approach offers the following advantages:

- Absence of high-spatial-frequency errors;
- Easy and fast to make many encoders once the setup is in place;
- Relatively easy to make an encoder with a pitch on the order of a wavelength;
- and
- A "fringe locker" can be used to stabilize the interference fringes for long exposures.

A major disadvantage of the standard holographic approach is the requirement for collimating optics whose diameters are on the order of the length of the desired encoder, and which require tight control of spherical aberration in order to maintain a uniform pitch for the entire length of the encoder.

We describe here a method for making encoders of uniform pitch without the use of collimating optics.

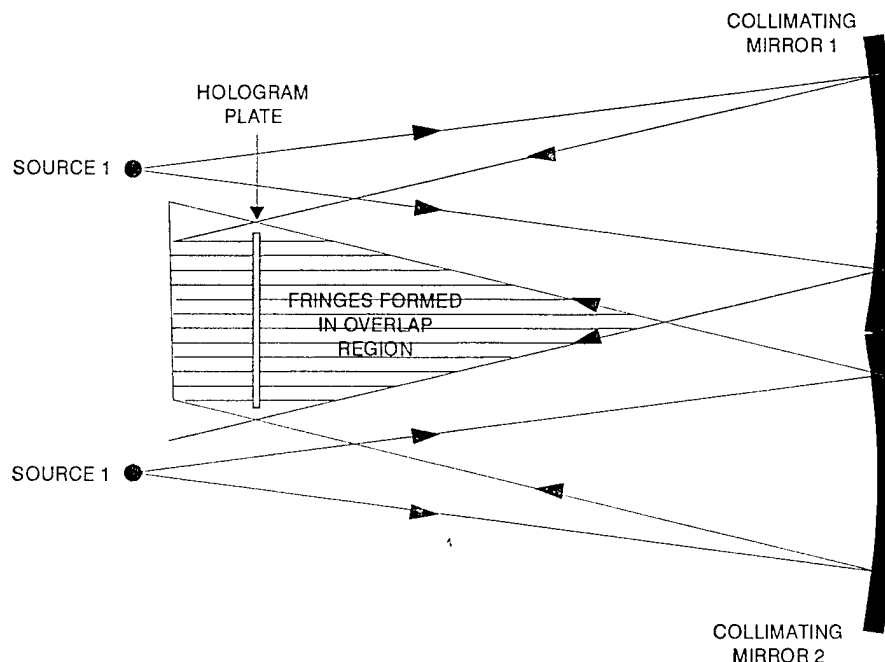


Figure 1 Making a Holographic Encoder with Collimating Optics

THE BASIC APPROACH

When interference from two point sources occurs, bright fringes are formed on surfaces which are the loci of points for which:

$$R_1 - R_2 = m\lambda, \quad (1)$$

Where R_1 and R_2 where the distances from the point on the surface to each of the two point sources, λ is the wavelength of the light, and m as an integer—the order of interference. Each surface of this type is a hyperboloid of revolution about an axis formed by the line joining the two point sources. These hyperboloids intersect a plane containing the two point sources in a series of hyperbolae given by:

$$(x/a)^2 - (y/b)^2 = 1, \quad (2)$$

Where, if $2A$ is the separation between the two point sources (located symmetrically on the x -axis), $a = m\lambda/2$ and $b = \sqrt{A^2 - a^2}$. Figure 2 shows these hyperbolae for a pair of point sources that are 20 wavelengths apart.

Our approach for making an error-free encoder without collimating optics is to find a smooth curve which, when overlaid on figure 2, intersects the family of hyperbolae at equal intervals (the encoder pitch) along the curve.

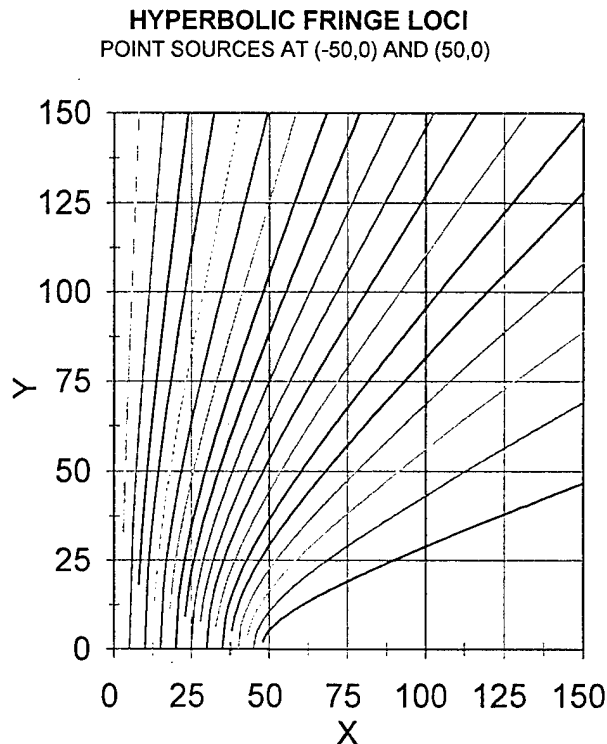


Figure 2 Hyperbolic Fringes from 2 Point Sources

The figure should be reflected about the X and Y axes to show the entire fringe pattern. The point sources are at $(-50,0)$ and $(50,0)$, and the wavelength is 5. The fringes intersect the X -axis at intervals of $\frac{1}{2}$ -wavelength.

Although we have not yet found an analytic solution to the problem of generating this curve, we have developed a numerical technique for generating families of such curves. The

numerical solution parallels a geometric solution that goes as follows. Starting with the hyperbolae shown in figure 2, take a pair of dividers and set the space between the points to the desired fringe spacing on the encoder. Next put one of the divider points on the Y-axis at a height where the lateral distance between the hyperbolae is a little less than the selected fringe spacing. Now, keeping the first point of the dividers where it is, rotate the dividers so that the other point falls on the next hyperbola to the right. There are two possible intersection points: pick the one closer to the X-axis. Continue the process, going from one hyperbola to the next. The series of points on successive hyperbolae define the shape of the encoder curve.

We found it time-saving to select a fringe spacing and wavelength that were both scaled up by the same factor (e.g. instead of an actual wavelength of $0.4416\mu\text{m}$ and a desired encoder pitch of $10\mu\text{m}$, we used a wavelength of $44.16\mu\text{m}$ and a fringe spacing of $1000\mu\text{m}$). This, in effect, let us work with every 100th fringe and reduce the number of computations proportionately—without any significant loss in accuracy.

To extend the curve on the other side of the y-axis, repeat the same process but select the intersection points that are further from the x-axis—this will continue the curve without any discontinuity at the y-axis. Figure 3 shows a typical numerical solution. Note that in general the curve comes to an end—at a point where the distance to the next hyperbola is greater than the selected encoder spacing. The curves are also somewhat counter-intuitive in that they are not symmetric about the y-axis.

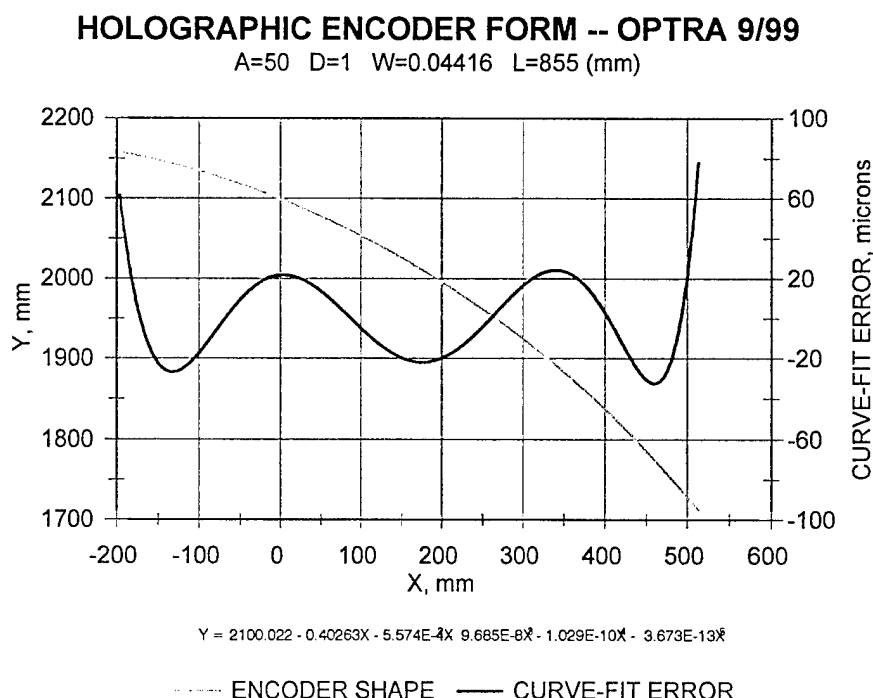


Figure 3 Numeric Solution to Error-Free Linear Encoder Problem

The blue line is the shape of the encoder and the red line is a 5th power polynomial fit to the numeric solution.

Making the encoder

Once the shape of the curve is defined and selected as described above, it must be fabricated. We are interested in encoders up to a meter long, and have selected the geometry shown in figure 4, in which the point sources are 100mm apart and the laser wavelength is 441.6nm (HeCd laser). This figure shows a pair of encoder shape solutions, either of which would serve our purposes.

Clearly we need two material items to proceed: (1) a rigid and thermally stable hologram substrate that can be formed into the desired shape, and (2) a form that can be used to define the shape of the encoder, and to hold it in place during exposure.

The form can be made in a straightforward manner from either the numeric solution to the problem, or from a polynomial best-fit to the numeric solution. Although the machining of the form should be smooth and accurate, small localized errors will probably have little effect.

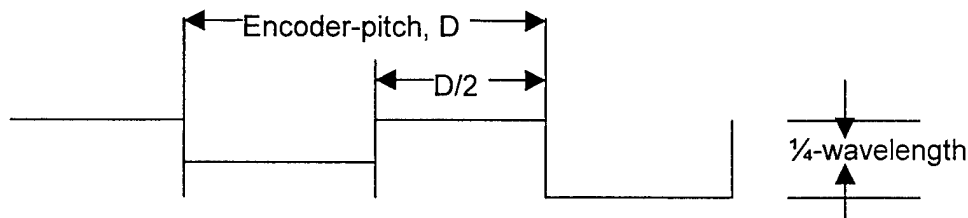
The very nature of bending the substrate to fit the form acts very effectively to prevent errors of high-spatial frequency (the substrate simply won't follow such errors). The thickness of the substrate will be selected on the basis of its yield strength, the curve to be followed, and the method used to hold the substrate in place on the form. It is not yet obvious (to us) whether a convex or concave form is preferable. Glasses are readily available in thickness that will allow them to be easily bent into the shapes that are needed.

Once the master hologram is exposed and processed to give the desired groove profile, it will be placed on a flat substrate and used in the manufacture of replicas—which can be on thicker glass or metal. Zerodur and Invar are clear choices if thermal stability is an issue.

We have used this process to make a test encoder against a circular form. The circular form does not, of course, produce a constant pitch over the length of the encoder, but it allowed us to validate the feasibility of the fabrication concept. The resultant encoder had just the pitch errors predicted by modeling.

Controlling the groove profile

Our specific goal is the manufacture of an encoder to be used in reflection. Ideally, when the encoder is illuminated normally by a small diameter beam of specified wavelength, there will be no light in either the zero order or in the ± 2 order—thus minimizing contamination of the ± 1 orders used to make position measurements. These two requirements are met by an encoder having a 50% duty cycle and a height of $\frac{1}{4}$ -wavelength (see figure 5). There are a variety of ways to manufacture this profile—through a combination of substrate preparation, exposure control, and post-exposure processing.



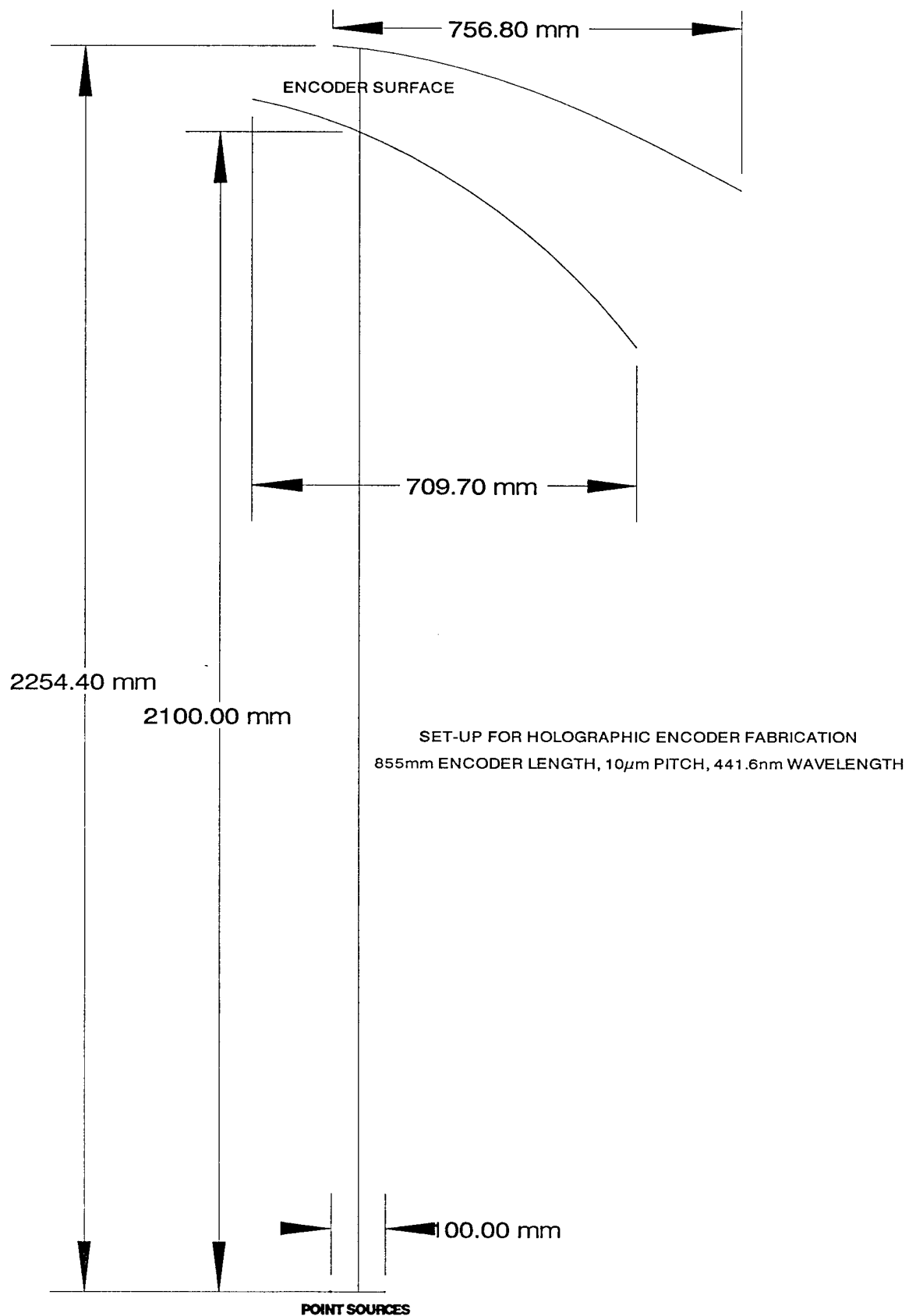


Figure 4 Overall layout for Holographic Encoder Fabrication
By using a precisely shaped encoder substrate, an error-free encoder can be made holographically without the need for collimating optics (mirrors or lenses):

Summary

A Wiener filter was developed to perform interpolation of Core grid data. An analysis was performed to determine the optimal spatial interval for interpolation of the Core data and number of filter coefficients that provides a peak-to-peak error of less than 20 nm. Linear interpolation was also performed in order to obtain a point of comparison. This memo describes the results of the analysis.

Core data

Figure 1 below shows a plot of the Core error data over a 2-mm range. The data was sampled at a 5- μm spatial interval. The peak-to-peak error of the uncalibrated Core data is 117.3 nm and the rms magnitude of the error is 28.4 nm. The goal of any interpolation scheme is to substantially reduce both of these error magnitudes.

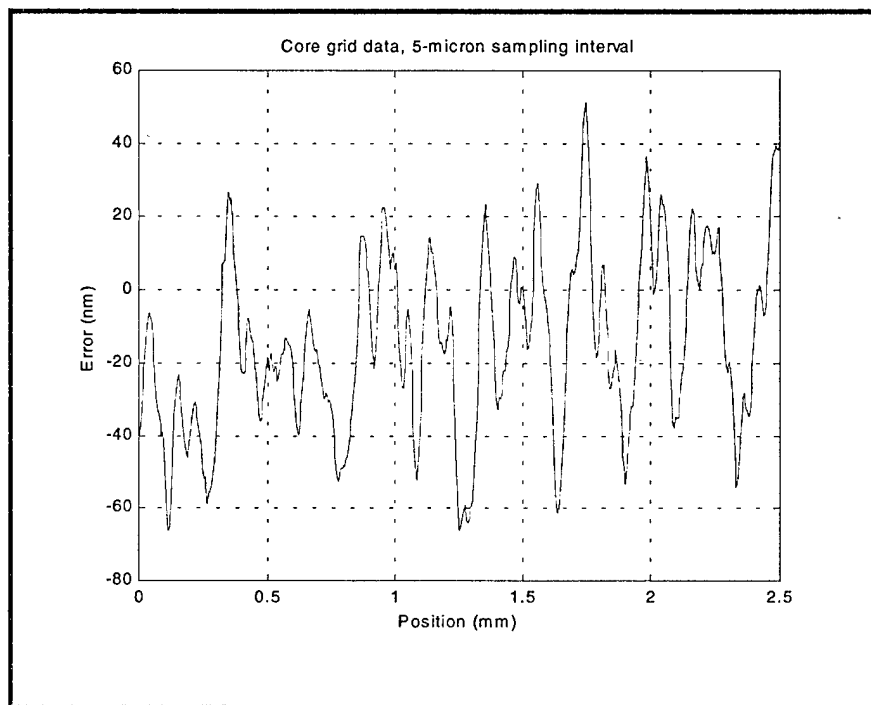


Figure 1. Core grid error data, 117.3 nm p-p error, 28.4 nm rms error

Wiener Filter

The goal of a Wiener filter is to perform optimal estimation of the data by minimizing the mean-square error. This is effectively done by appropriate weighting of the available data, which are then used to arrive at an estimate. The weighting for a given data value is obtained by the autocorrelation function of the data. Figure 2 below shows two plots of the autocorrelation function of the data, the upper is shown over the whole range and the lower is a blow-up of the first 500 microns.

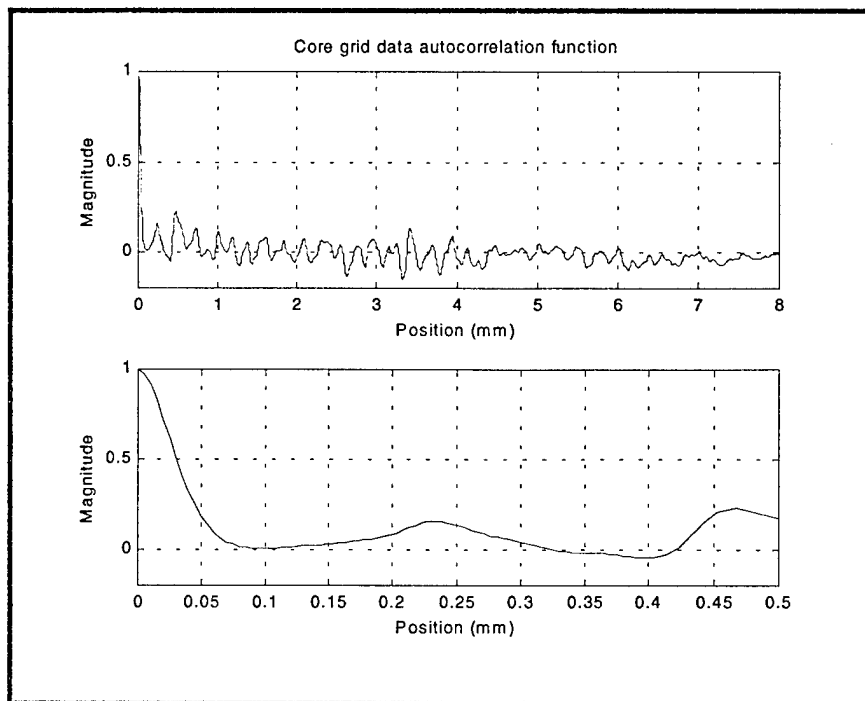


Figure 2. Core grid data autocorrelation function

The lower figure shows that the correlation between points quickly rolls off to zero within a spatial interval of 100 microns. This is an expected result, as 100 microns is the spot size of the laser on the grid. The upper figure also shows that there appears to be some cyclic response to the data as various peaks occur for the first 4-5 mm. This is somewhat evident in the data itself, as from figure 1 there appears to be some evidence of different frequency components to the error. However, due to noise or other processes, the magnitude is never high enough beyond 100-microns where one would expect a large, meaningful contribution from these data.

So, how then does a Wiener filter use the autocorrelation data to provide an estimate? Say that one wanted to set up a calibration table with the data points separated by 0.1 mm intervals. Now, take for example a situation where a command puts you between two points in the table, with a separation between one point of 0.01 mm and the other of 0.09 mm. Based on the autocorrelation data, the Wiener filter will more heavily weight the data point with higher correlation (the one that is only 0.01 mm away) when generating the estimate, which may result in a coefficient of 0.9 for the closer point and 0.1 for the further point. In contrast, linear interpolation weights each datapoint equally, e.g. each datapoint would be multiplied 0.5 and then the sum is taken to generate the estimate. Furthermore, the Wiener filter can be expanded to include more and more neighboring points to possibly improve the estimate.

Core Data Analysis

The Core data was analyzed to determine the effect of increasing the spatial separation between points on the calibrated peak-to-peak error and rms error. This was done by the following manner.

1. A 2.5 mm swath of data was analyzed.
2. The data was down-sampled to obtain the appropriate separation between data points.
3. The data was then interpolated using both a Wiener filter and linear interpolation to obtain estimates at the missing points.

4. An error signal was obtained as the difference between the original data and the interpolated data.
5. The peak-to-peak error was obtained from the maximum and minimum values of the error signal.
6. The rms error was obtained from the error signal.
7. The process was repeated for different sampling intervals and the number of datapoints to use to obtain the estimate, e.g. 2 is the neighboring 2 datapoints, ideally 1 on either side, 4 would be the neighboring 4, ideally 2 on either side, and so on. A 2-datapoint estimate results in a Wiener filter with two coefficients, a 4-datapoint estimate results in a Wiener filter with 4 coefficients, etc.

The table below shows the relationship between downsampling amount and number of datapoints needed for a 300-mm grid.

Core grid table calibration requirements			
Downsampling amount	Spatial interval between points (μm)	Number of points for single axis (300-mm)	Number of points for grid (300-mm)
2×	10	30000	900×10^6
4×	20	15000	225×10^6
6×	30	10000	100×10^6
8×	40	7500	56×10^6
10×	50	6000	36×10^6

The goal of any interpolation scheme is to increase the sampling interval, thus decreasing the number of points needed for storage by the servo controller or other piece of hardware used to perform the calibration.

Results

The following tables show the results for the Core data and Figure 3 shows a plot of the case using a 4-point Wiener filter with a sampling interval of 20 μm . The upper figure shows an overlay of the Core data with the Wiener estimate and the estimate obtained from linear interpolation and the lower figure shows the error functions. The improvement for this case over the uncalibrated error is 10× in both peak-to-peak error and rms error.

Core Grid Calibration Peak-to-Peak Error (nm)					
Downsampling Amount	Linear Interpolation	Wiener Estimate			
		2 points	4 points	8 points	16 points
2×	9.2	8.2	7.2	6.4	6.6
4×	31.6	18.6	15.0	15.7	15.6
6×	45.9	32.7	29.3	30.8	30.9
8×	60.9	55.1	54.6	54.5	55.1
10×	70.7	66.4	69.1	68.2	72.0

Core Grid Calibration RMS Error (nm)					
Downsampling Amount	Linear Interpolation	Wiener Estimate			
		2 points	4 points	8 points	16 points
2×	1.2	1.2	1.0	0.96	0.91
4×	4.7	3.1	2.3	2.1	2.1
6×	8.1	5.4	4.8	4.7	4.7
8×	11.2	8.3	8.1	8.0	8.1

10×	14.0	10.7	10.5	10.5	10.5
-----	------	------	------	------	------

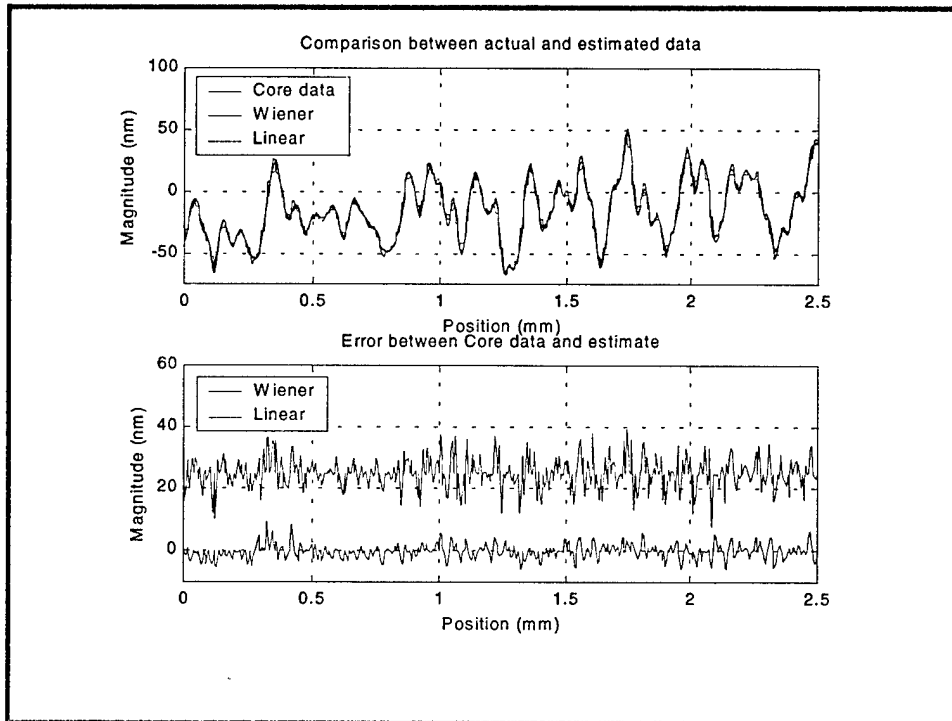


Figure 3. Data for 4× (20- μ m interval) case using 4-point Wiener estimate (Linear estimate offset by 25 nm bottom chart for clarity)

Conclusion

An analysis was performed on Core grid data to determine the improvement that could be obtained using a Wiener filter to interpolate the data. The analysis indicates that the high spatial frequency profile of the error requires a sampling interval no greater than 20 μ m to obtain a peak-to-peak error less than 15-nm and an rms error less than 2.5 nm. While the improvement in performance is substantial at 10× better than the uncalibrated error of 117 nm peak-to-peak and 28 nm rms, the practicality of implementation is most likely low as it would require a 225 million point error table for a 300-mm grid.